Common genomic pathways between endometriosis and fibromyalgia

Thesis submitted for the degree of Master's of Science by Research

Ioana-Wilhelmina Lucinescu

Lady Margaret Hall

University of Oxford

Word count (Total): 15,525

Word count (Abstract): 296

Supervised by: Prof Katy Vincent, Prof Krina Zondervan, Dr Nilufer Rahmioglu, Dr Anne Ndungu

Table of contents

Abstract 3

Acknowledgments 5

List of figures 6

List of tables 9

List of abbreviations 11

Chapter I. Introduction and Aims of thesis 12

1.1. Introduction 12

1.2. Aims 23

Chapter II. Systematic review of genome-wide association studies of primary pain conditions 24

2.1. Rationale of systematic review 24

2.2. Methods 25

2.3. Results 30

2.4. Discussion 35

Chapter III: Genetic analysis of fibromyalgia in UK Biobank 38

3.1. Rationale 38

3.2. Methods 38

3.3. Results 45

3.4. Discussion 60

Chapter IV. Genetic sharing of endometriosis and fibromyalgia 64

4.1. Introduction 64

4.2. Methods 65

4.3. Results 68

4.4. Discussion 79

Chapter V. Discussion and Conclusions 83

5.1. Discussion 83

5.2. Conclusions 90

References 91

Abstract

Endometriosis and fibromyalgia are often comorbid conditions and are part of a larger group of conditions known as chronic overlapping pain conditions which share many symptoms and pain mechanisms. Endometriosis is a secondary chronic pain condition, where there is peripheral pathology underlying the pain, whereas fibromyalgia is a primary chronic pain condition, with no known underlying pathology.

The aim of my thesis is to uncover the shared genetic basis of fibromyalgia and endometriosis. To investigate current knowledge of the genetic overlap between primary and secondary chronic pain conditions, I conducted a systematic review. I identified *NGF*, *ASTN2* and *LRP1* as possible pain susceptibility genes with pleiotropic effects in both primary and secondary pain conditions. To address the knowledge gap in the genetics of fibromyalgia, I conducted a genome-wide association study (GWAS) in UK Biobank, which illustrated that different case definitions show genetic heterogeneity. To boost power of discovery, I meta-analysed UK Biobank ICD-based fibromyalgia GWAS with a fibromyalgia GWAS from FinnGen. One lead SNP was identified, rs34323745, which nearly reached genome-wide significance ($p = 7.269 \times 10^{-8}$). To study the shared genetic basis of endometriosis and fibromyalgia, I conducted a multi-trait association study using the meta-analysis of fibromyalgia from UK Biobank and FinnGen and the largest published meta-analysis of endometriosis to date (Rahmioglu *et al*., 2023). I identified two genome-wide significant hits, rs17082358 on chromosome 2 in the *SYNE1* locus ($p = 1.49 \times 10^{-11}$) and rs13432756 on chromosome 6 in the *GREB1* locus ($p = 6.39 \times 10^{-9}$). Both loci have previously been associated with endometriosis. *GREB1* was previously associated with fibromyalgia, but not in a large genetic study. Future studies are needed to replicate these findings and to

better understand the biology of *GREB1* and *SYNE1* as shared genes in fibromyalgia and endometriosis.

Acknowledgments

I would like to firstly thank my supervisors, Prof Katy Vincent, Prof Krina Zondervan, Dr Nilufer Rahmioglu and Dr Anne Ndungu, for their dedication to helping me reach my goal and for their valuable insight and wisdom. I would like to take a moment to remember Dr Anne Ndungu, which had just started contributing to my supervision when she unfortunately passed away. Next, I would like to thank Medical Research Council, Eli Lily and Nuffield Department for Women's and Reproductive Health for supporting me throughout this wonderful project and for their generous financial support. I would also like to thank Dr Nina Shigesi, Dr Danielle Perro and Dr Lydia Coxon for their indispensable guidance and help throughout my project.

List of figures

Figure 1. PRISMA flow-chart of the systematic review. 31

Figure 2. Chord plot of variants of interest (top, rsids) and the traits they are associated with according to Phenoscanner[71,72]. 33

Figure 3. Heatmap of genes of interest in 53 tissue types from GTEx version 8. Figure was produced by FUMA[73]. 34

Figure 4. Sex-stratified Miami Plot of fibromyalgia in UK Biobank. Female only GWAS results given as a Manhattan plot on the top and male only GWAS results given as a Manhattan plot on the bottom. Each dot on the plot is a SNP and significance of association is given on Y-axis. Red-dashed line represents genome-wide significance threshold ($p<5x10^{-8}$), and the blue line represents a relaxed threshold for nominal significance ($p<5x10^{-5}$). 49

Figure 5. Manhattan plots of clinical diagnosis of fibromyalgia GWAS using different methods: BOLT-LMM (A), REGENIE (B) and SAIGE (C). Blue line suggests a nominal threshold of significance ($p<10^{-5}$). 51

Figure 6. Miami plot of sensitivity analysis between clinical-confirmed fibromyalgia GWAS (top, labelled Hospital), and self-reported GWAS (bottom) 53

Figure 7. Q-Q plots of clinical-confirmed fibromyalgia GWAS (left) and self-reported fibromyalgia GWAS (right). 54

Figure 8. Miami plot of ICD-based fibromyalgia GWAS (top) versus primary care fibromyalgia GWAS (bottom). The blue line indicates suggestive level of significance ($p<10^{-5}$) 55

Figure 9. Q-Q plots of ICD-based fibromyalgia GWAS (left) and primary care fibromyalgia GWAS (right). 56

Figure 10. Manhattan plot for meta-analysis of fibromyalgia. Blue line represents a nominal threshold of significance ($p<10^{-5}$). Red line represents genome-wide significance ($p<5x10^{-8}$). 58

Figure 11. Miami plot of GWAS meta-analysis of fibromyalgia (top) versus primary care GWAS (bottom). Blue line represents a nominal threshold of significance ($p<10^{-5}$). Red, dotted line represents genome-wide significance ($p<5x10^{-8}$). 60

Figure 12. Q-Q plot of multi trait association analysis of fibromyalgia. 69

Figure 13. Manhattan plot of multi trait association analysis of fibromyalgia. Dotted line represents genome wide significance ($p<5x10^{-8}$). 70

Figure 14. Regional association plots of SYNE1/6q25.1 (I) and GREB1/2p25.1 (II) loci, the top signals identified from the multi trait association analysis (MTAG) of fibromyalgia and endometriosis (purple dot). A: Endometriosis MTAG; B: Endometriosis meta-analysis; C: Fibromyalgia MTAG; D: Fibromyalgia meta-analysis. 72

Figure 15. Heatmap of genes expression for genes of interest in 54 tissue types from GTEx v8(56). Lead variants ($p < 10^{-5}$) from multi trait association analysis were mapped positionally to genes by FUMA. Figure was produced from FUMA. 74

Figure 16. Heatmap of gene expression for genes of interest in 54 tissues from GTEx v8(56). Lead variants ($p < 10^{-5}$) from multi trait association analysis have been mapped based on expression quantitative trait loci from BRAINEAC and eQTLGen databases(100,101). Figure was produced by FUMA. 75

Figure 17. Gene enrichment analysis of GWAS reported genes. The y axis represents gene sets with adjusted P-value < 0.05. Figure was produced by FUMA. 76

Figure 18. Differentially expressed gene sets analysis. The gene sets are determined by selecting all significant genes, with a Bonferroni corrected p-value < 0.05 and absolute log fold change >= 0.58, from a two-sided t test per tissue versus remaining tissues. Red bars represent significantly enriched tissues with Bonferroni corrected p-values (p < 0.05). Figure was produced by FUMA. 77

List of tables

Table 1. List of primary pain conditions for which genome-wide association studies have been performed and are available in GWAS Catalog[67]. 26

Table 2. List of variants in linkage disequilibrium associated with different primary pain traits ($r^2 > 0.8$). 32

Table 3. Mapped genes for the variants of interest with the FUMA tool with SNP2GENE option[73]. 34

Table 4. Fibromyalgia case and control definitions in UK Biobank for genome-wide association study. 40

Table 5. Sociodemographic descriptives of fibromyalgia cases in UK Biobank by case ascertainment category. 47

Table 6. Top lead SNPs ($p < 10^{-7}$) of female-only, male-only, and sex-combined genome-wide association studies of fibromyalgia in UK Biobank. 50

Table 7. Independent lead SNPs ($p<10^{-7}$) from the assessment of three GWAS methods using clinical fibromyalgia UK Biobank cases. EA, effect allele. Missing values for several SNPs in SAIGE columns are due to them failing extra QC that SAIGE employs. 52

Table 8. Top 15 lead SNPs (p < 5 x 10-6) of clinical fibromyalgia GWAS and self-reported fibromyalgia GWAS. EA, effect allele. 54

Table 9. Lead SNPs ($p<10^{-6}$) of fibromyalgia primary care GWAS, ICD GWAS and FinnGen GWAS. 57

Table 10. Expression quantitative trait loci for the rs34323745, the top GWAS signal in the fibromyalgia GWAS meta-analysis. Results were obtained from GTEx Portal (v8)[75]. 59

Table 11. Expression quantitative trait loci for rs34323745, the lead GWAS signal in the fibromyalgia GWAS meta-analysis. Results were obtained from FUMA, using eQTLGen and BRAINEAC databases. False Discovery Rate (FDR) was calculated by FUMA. 59

Table 12. Lead SNPs (p < 10-5) from multi-trait association analysis of fibromyalgia (MTAG) versus univariate GWAS of endometriosis (ENDO) and fibromyalgia (meta-analysis, UK Biobank and FinnGen). EA, effect allele; EAF, effect allele frequency. 71

List of abbreviations

COPC	Chronic overlapping pain conditions
GWAS	Genome-wide association study
SNP	Single nucleotide polymorphism
IBD	Inflammatory bowel disease
RA	Rheumatoid arthritis
eQTL	Expression quantitative trait locus
MAF	Minor Allele Frequency
FDR	False Discovery Rate
maxFDR	Maximum False Discovery Rate
chr	chromosome

Chapter I. Introduction and Aims of Thesis

1.1. Introduction

1.1.1. Chronic pain

Chronic pain is defined as pain that recurs or persists for more than three months. It is a burdensome symptom that affects 1 in 3 people in the UK (Chronic pain in adults 2017, Health Survey for England) with a significant impact on quality of life[1]. Some patients are thought to be more at risk of developing chronic pain, known as pain vulnerability[2]. There are many factors that are thought to underlie chronic pain risk, such as genetics, sex, stress, personality, and others. Twin studies have shown chronic pain is heritable[3–5].

1.1.2. Classification of chronic pain conditions

According to the International Classification of Diseases 11th Revision (ICD-11), chronic pain can be classified as primary or secondary, depending on whether the pain can be attributed to another condition[6]. Primary chronic pain is a category of conditions that have no known underlying cause, and pain is the main symptom. Primary pain conditions are classified depending on location of symptoms as widespread, visceral, orofacial, and musculoskeletal. Prevalence is estimated to be between 1% and 6% in the UK[6,7].

Secondary chronic pain is a group of conditions where pain can be attributed to an underlying diagnosis[6]. According to ICD-11, they are grouped based on location of pain as musculoskeletal, visceral, orofacial, or based on origin of pain as post-surgical, post-traumatic, cancer and neuropathic[6]. Examples of secondary pain conditions include endometriosis, inflammatory bowel disease (IBD), osteoarthritis and rheumatoid arthritis (RA). These conditions are common, affecting many people,

and complex, the mechanism of action involving multiple biological pathways and gene-environment interactions. Although pain is a commonly reported symptom in patients with common, complex diseases, not all patients will develop chronic pain alongside their condition. This is thought to be due to pain vulnerability: some patients are more prone to developing chronic pain than others with the same condition[2].

1.1.3. Mechanisms of chronic pain

Proposed mechanisms of chronic pain include damage to neural tissue (neuropathic pain), damage to non-neural tissue (nociceptive pain) or sensitisation either of the central or peripheral nervous system to nociceptive stimuli (nociplastic pain). Patients can experience mixed pain types. Depending on pain mechanisms, there are different pharmacological treatment options. Neuronal pathways implicated in nociception are neuronal transmission of painful stimuli from peripheral tissue to the spine, descending pain modulatory pathways that inhibit or encourage pain transmission from the spine to the brain, and lastly, within central nervous system communication that regulates emotional or cognitive responses to pain stimuli[8–10].

The pain mechanism in primary chronic pain conditions is thought to be nociplastic[11,12]. Nociplastic pain arises due to a change in pain perception or a change in the nociceptive pathways, without any evidence of peripheral pathology[13]. Common mechanisms include peripheral sensitisation, use-dependent synaptic plasticity of spinal neurons, thalamocortical mechanisms of attentional control, expectation, and decreased descending inhibition[14]. In contrast, chronic secondary pain mechanisms are largely neuropathic or nociceptive[13]. Neuropathic pain causes include inflammation and auto-immunity, which are present in complex, common

conditions, such as endometriosis, IBD and RA[14]. An example of nociceptive secondary pain condition is osteoarthritis[6].

Chronic pain patients experience a reversible decrease in gray matter in different brain areas. A non-exhaustive list of brain areas that decrease in gray matter volume following chronic pain onset is bilateral medial frontal gyri, bilateral superior frontal gyri, right pre- and post- central gyri (including the S1 and primary motor cortex), bilateral insula (anterior), right cingulate cortex (dorsal posterior cingulate cortex), basal ganglia, thalamus and periaqueductal gray[15]. Different studies have found additional brain areas that are active depending on different aspects of pain, such as anticipation, attention, placebo effect, mood disorders, like anxiety and depression and pain catastrophising[16–20]. Descending pathways that facilitate pain are thought to fire more in chronic patients[21–23]. The prefrontal cortex is both an important area for emotional regulation of pain and for controlling interactions between brain areas involved in nociception. Neurodegeneration of the prefrontal cortex is noted in many chronic pain patients, and this is thought to result in decreased inhibition of the central nociceptive pathways[24].

1.1.4. Genetics of chronic pain

Chronic pain as a trait shows moderate to strong genetic influence[25–27]. Identifying which genetic risk factors predispose people to develop chronic pain has been researched primarily using linkage analysis and association studies. Researchers also used animal models to study the inheritance of pain and analgesic sensitivity in different mice strains and model the development of chronic pain[2,28].

The basis of linkage studies is genes being transmitted together from generation to generation in a family. The degree of linkage is based on physical distance on the

chromosome: genes which are physically closer tend to be inherited together. Many gene pairs show an incomplete level of linkage, meaning recombination happens at a constant rate between them. Linkage analysis involves studying the transmission of alleles of polymorphic genes from typed individuals of a family. Alleles are typed throughout the genome of affected and unaffected individuals from a pedigree to investigate the segregation between marker and disease. Association between marker and disease pinpoints a specific area for further study[29]. Linkage analysis studies are most useful when examining single gene disorders because the mode of inheritance is known, and there is a clear relationship between genotype and phenotype. Researchers have looked at rare, single gene pain disorders that result in insensitivity to pain, known as allodynia. Examples include hereditary sensory and autonomic neuropathies, the familial hemiplegic migraine disorders, and neurological channelopathies presenting as paroxysmal pain disorders[2,28]. Linkage studies have identified mutations in these patients which impact proteins such as ion channels (*KCNS1*), neurotrophic factors (*NGFB*), enzymes (*COMT*), and receptors (*OPRM1*, *5HTR2A*)[8,28,29].

There are many genetic and environmental factors that play a part in pathogenesis of chronic pain condition[30]. Having two copies of an allele at one locus is no longer enough to cause disease, unlike in the case of monogenic disorders. While having a family member with chronic pain certainly increases risk of developing chronic pain, most cases are sporadic. It is thought chronic pain heritability is in part due to common polymorphic sites, also known as variants or single nucleotide polymorphisms (SNPs), which accumulate and increase the risk to develop the condition[28,30]. Individually, common variants (with a frequency of more than 1%) are thought to have only a modest contribution to increasing risk of disease (Odds ratios

from 1.05-1.2)[31]. This follows because if the effect of the variant would be detrimental to health, it would be selected against, which would eliminate it from the population. However, many causal variants are thought to be pleiotropic, meaning they are involved in multiple conditions. Complex diseases are also thought to be polygenic, meaning the risk from carrying different, common alleles that are associated with disease stacks up in additive fashion. The percentage of heritability due to common variants is usually low and there are other genetic risk factors that must be considered to explain the whole portion of genetic variability of complex diseases. One hypothesis is these common variants contribute to disease by acting as modifiers to other, rare variants[30]. Distinct, highly penetrant rare variants have been identified for many complex diseases, such as psychiatric disorders, due to exome and whole genome sequencing[2,30]. Nonetheless, common genetic variants are thought to play an important role in causing disease, however, often the case number is not great enough to detect associations[32].

Since linkage analysis studies have limited use in complex diseases, association studies are now the common tool being used. The basis of association studies is linkage disequilibrium. This phenomenon occurs when alleles within a haplotype are inherited together more often than they are not, due to recombination. Association studies were often used to fine map a region of interest thought to be causal to disease[33]. This region is mapped in both cases and controls, and allele frequencies are compared using chi-squared test or Fisher's exact test[34]. These candidate gene studies are often subject to bias, most obvious being selecting which region of interest to analyse, rather than looking at the whole genome. They also suffer from a lack of reproducibility. There are thought to be high rates of type I and type II errors in these studies[33].

Genome-wide association studies (GWAS) are a useful approach for studying common, complex diseases, including chronic pain. They involve a genome-wide scan of a panel of known SNPs, in a large number of cases and controls in order to identify common variants that predispose to developing a disease. Information on potential causal variants that are not measured directly can be recovered by imputation. Imputation is an inference based on nearby SNP and haplotypes derived from a fully sequenced reference panel[35].

GWAS have uncovered many genetic variants which provided useful insights into the biology and heritability of chronic pain conditions and have offered suitable treatment targets[36]. However, to date, none of the GWAS in complex conditions like endometriosis has focused on chronic pain.

Large GWAS were conducted into both primary and secondary pain conditions. There have been numerous genes associated with chronic pain, particularly for musculoskeletal conditions and migraine, that affect a range of biological pathways, including neurotransmission, immune function, metabolism, protein degradation, cellular growth, and others. Limitations of GWAS include clinical definitions of chronic pain phenotypes; lack of reproducibility, especially in animal models; and study size[28,36,37]. Several neuroimaging studies have successfully linked brain changes during chronic pain with genetics, by looking at catecholamine and serotonergic pathways[8].

GWAS can point to new drug targets and improve our understanding of chronic pain mechanisms, but they also lead to many loci being associated to disease, without shedding light into the causal relationship between common variants and disease. Novel methods can now be used to infer causal relationships. A recent study on the

shared genetics of chronic pain conditions showed, using latent causal variable analysis, that psychological traits have a genetic causal effect on chronic pain traits and vice versa[38]. Depression and anxiety-related traits increased the risk of chronic back, neck/shoulder, hip, knee, abdominal, headache and widespread pain, whereas chronic back pain, facial pain and headaches had genetic causal effects upon depression and anxiety[38]. Mendelian randomisation is another method which can be used to infer causality[39]. As an example, in the latest meta-analysis of endometriosis, MR was used in conjunction with methylation and expression quantitative trait loci information to identify genes directly causal to endometriosis[40].

1.1.5. Fibromyalgia

Fibromyalgia syndrome is a chronic primary pain condition, according to ICD-11[6]. It is characterised by diffuse musculoskeletal pain in at least 4 of 5 body regions and in at least 3 or more body quadrants (as defined by upper–lower/left–right side of the body) and is associated with sleep and affective disturbances. The nature of the pain is nociplastic, meaning a lack of underlying pathology in the affected tissues, and spontaneous or evoked pain, accompanied by allodynia and/or hyperalgesia[41,42]. Current diagnoses criteria used are the American College of Rheumatology (ACR) 2010 (revisited in 2016) and a multidimensional fibromyalgia diagnosis framework created by the collaboration between Analgesic, Anesthetic, and Addiction Clinical Trial Translations Innovations Opportunities and Networks public-private partnership with the U.S. Food and Drug Administration and the American Pain Society, known as AAPT criteria, published in 2018. They largely agree, with the AAPT being more controversial. They both evaluate the generalisability of pain, sleep disturbances and fatigue, using different criteria[43–45]. The ACR 2016 updated criteria are:

(1) The presence of widespread pain, defined as pain in at least 4 out of 5 regions.
(2) Symptoms have been present at a similar level for at least 3 months.
(3) Widespread pain index (WPI) ≥ 7 and symptom severity scale (SSS) score ≥ 5 or a WPI of 4 - 6 and SSS score ≥ 9.
(4) A diagnosis of fibromyalgia does not exclude the presence of other clinical diagnosis(23).

Prevalence of fibromyalgia is around 2% worldwide, but estimates vary across the globe and depending on diagnostic criteria used[46]. It is widely recognised as predominantly a female syndrome; however, recent unbiased studies revealed a less drastic, 1.5:1 female to male ratio. Age at diagnosis is between 50 and 60 years old, however this is contested, with patients of all ages now receiving a diagnosis of fibromyalgia[47].

Comorbidities often include regional pain syndromes, such as irritable bowel syndrome, dysmenorrhea, chronic headaches, and mental health disorders, like anxiety and depression, but also adverse early-life events and post-traumatic stress disorder[43].

Patients with fibromyalgia are thought to have an imbalance between nociceptive and anti-nociceptive signalling pathways, with a decrease in descending inhibitory pathways, coupled with an increase in excitatory neurotransmitters in brain areas related to pain, as well as decreased threshold for pain in the peripheral tissues. Neuroinflammation and small fibre neuropathy are thought to be the drivers behind peripheral sensitisation[43,48–50].

1.1.6. Genetics of fibromyalgia

Heritability of fibromyalgia is estimated to be ~50%, with a clear pattern of increased familial risk ratios. The first-degree relatives of fibromyalgia patients are eight times more likely to have fibromyalgia compared to family members of healthy controls, and they also report very high rates of other chronic pain conditions[12].

Genetic polymorphisms were identified in serotonergic (*5-HTT*, *COMT*, *SLC64A4*, *HTR2A*); catecholaminergic (Dopamine D4 receptor, μ1 opioid receptor, *RGS4* and β-2 adrenergic receptor); cannabinoid (*CNR1*) and GABA (*GTPCH* A receptor, *GABRB3*) signalling pathways, as well as ion channels involved in nociception (*TRPV2*, *SCN9A*)[43,51,52].

The largest published GWAS for fibromyalgia has less than 1000 cases. The authors did not find any genome-wide significant associations ($p<5*10^{-8}$), however, they identified a copy number variation, that is, a deletion or duplication of a section of a chromosome, in *NRXN3*, a neuronal cell surface protein involved in cell-cell adhesion, and a nominal association for *MYTL1*, a neuronal specific transcription factor, which was replicated in another cohort[53].

Linkage studies confirmed a strong genetic component in fibromyalgia, and suggested linkage of fibromyalgia to the chromosome 17p11.2-q11.2 region[54]. A large candidate gene study by Smith *et al*. (2012) identified three genes potentially involved in fibromyalgia mechanisms, *GABRB3* (rs4906902, $p = 3.65 \times 10^{-6}$), *TAAR1* (rs8192619, $p = 1.11 \times 10^{-5}$); and *GBP1* (rs7911, $p = 1.06 \times 10^{-4}$). The association between *TAAR1* and fibromyalgia was replicated using a second independent cohort of fibromyalgia patients[55]. T*AAR1* is an intracellular G-protein coupled receptor, which binds to trace amines and is expressed in low amounts in several different

brain regions, namely pituitary, prefrontal cortex, hypothalamus, hippocampus, amygdala, dorsal raphe nucleus and ventral tegmental area. It is also expressed in immune cells and thyroid. It interacts with other monoaminergic systems in the brain to either potentiate or inhibit the response of neurons to monoamines, in a brain region dependent manner[56]. It is thought to play a role in increasing sensitivity to pain[57]. Its role in chronic pain may also be connected to its function in the reward system[56].

1.1.7. Endometriosis

Endometriosis is a common, complex disorder, in which tissue resembling the endometrium grows outside of the uterus, mainly in the pelvic area including the ovaries, ligaments and peritoneal surfaces as well as the bowel and bladder[58].

Histological presentation varies from superficial peritoneal lesions to endometriosis cysts and nodules and can be accompanied by scarring and adhesions. Importantly, endometriosis associated pain is not correlated with disease severity[58]. Women with endometriosis present with generalized pain in the pelvis and bowel, indicating a widespread visceral hypersensitivity and pain independent of lesion location[59].

The current understanding of pain mechanisms in endometriosis includes a pro-inflammatory environment caused by retrograde menstruation in the peritoneal cavity. Pro-inflammatory cytokines and chemokines, such as *PGE2*, *NGF*, *IL-8*, *IL1β*, *CCL5* and *TNFα* are elevated in the peritoneal fluid of patients. Long-term exposure to a pro-inflammatory environment can lower the threshold of activation of peripheral afferents, leading to peripheral sensitisation[59].

Brains of endometriosis patients present with a loss in gray matter volume in areas thought to be important for pain processing, such as thalamus or insula[58].

Endometriosis has a strong genetic component (heritability = ~50%)[25]. The largest meta-analysis of GWAS includes 22 studies and identified 42 genome-wide significant loci associated with endometriosis. Causal links were identified between associated variants and endometriosis, with pathways implicated in hormone signalling, uterine development, oncogenesis, inflammatory cell adhesion and angiogenesis[40]. Although chronic pain is a common symptom of endometriosis, not all patients experience pain, and even amongst those who do, experiences are varied[60]. Therefore, identifying which of these loci is associated with pain or with another process in the pathophysiology of endometriosis is not straightforward.

1.1.8. Chronic overlapping pain conditions

Chronic pain conditions that are commonly co-morbid are called chronic overlapping pain conditions (COPCs). This concept has been recognized by the National Institutes of Health and was initially highlighted in an Institute of Medicine report on “Relieving Pain in America”. The report acknowledged that some common or highly prevalent chronic pain conditions appear to coexist. These conditions are also more likely to affect women[61].

The list of recognised COPCs includes fibromyalgia, temporomandibular joint disorder, interstitial cystitis/painful bladder syndrome, irritable bowel syndrome, endometriosis, vulvodynia, chronic low back pain, chronic migraine, and chronic tension-type headache[62].

Twin studies suggest that approximately 50% of the risk of developing these conditions is genetic, and 50% is environmental[63].

In addition to pain, other common symptoms include fatigue, sleep impairment, problems with cognition, physical dysfunction, disturbances in affect (e.g., anxiety,

anger, depression). They more often affect women and are thought to have shared biopsychosocial mechanisms. Mechanisms of pain amplification or hyperalgesia seen in COPCs are thought to be the result of either peripheral and/or central sensitisation. These conditions appear to be related to such an extent that patients share more clinical signs and symptoms across pain conditions than within a specific pain condition[62,64]. This has prompted researchers and clinicians to think about COPCs as one condition, where pain manifests in different bodily regions during a patient's life[7]. Although there is increased prevalence of fibromyalgia in endometriosis patients, the mechanisms shared between these COPCs have not been identified[65,66].

1.2. Aims

The aim of this thesis is to identify the shared genetics of endometriosis and fibromyalgia.

In Chapter II, I conduct a systematic review of the current literature on genome-wide association studies in primary pain conditions and identify shared variants with secondary pain conditions. Fibromyalgia is a primary musculoskeletal pain condition and endometriosis is a secondary visceral pain condition. Genome-wide association studies have not been conducted specifically for pain, which makes identifying pain-related genes difficult. By focusing on primary pain conditions and then overlapping common variants with secondary pain conditions, I gathered evidence for the involvement of identified variants in chronic pain mechanisms. This is particularly relevant as for many of these variants, their role in chronic pain is unknown.

In Chapter III, I conduct a GWAS of fibromyalgia using UK Biobank cases and controls and then meta-analyse results with FinnGen GWAS. This is the largest

meta-analysis of fibromyalgia GWAS to date, which will contribute to furthering the knowledge of the genetic basis of fibromyalgia.

In Chapter IV, I focus on exploring the shared genetics of fibromyalgia and endometriosis using genetic correlation, multi-trait association analysis and functional annotation of identified shared variants. This will shed light on shared pain mechanisms between these conditions.

Chapter II. Systematic review of genome-wide association studies of primary pain conditions

2.1. Rationale of systematic review

Patients with complex diseases, such as endometriosis, rheumatoid arthritis, and inflammatory bowel disease often report pain alongside their condition[6]. Presently, no genome-wide association study has looked specifically at genetic signals underlying pain in these patients. However, many GWAS have been conducted for these conditions and have subsequently returned a plethora of genome-wide significant SNPs[28]. Current evidence suggests some patients are thought to be more at risk of developing chronic pain alongside these conditions[2]. However, due to the complexity of these diseases, uncovering which genetic variants contribute to pain vulnerability is difficult. To this end, I systematically reviewed the evidence that genetic variants associated with more than one primary chronic pain condition were also associated with endometriosis, osteoarthritis, RA and IBD, using publicly available GWAS results. The hypothesis tested in this review is that there is a common framework of pain genes impacted by common variants, which predisposes people to pain, regardless of tissue of origin of the chronic pain. Another hypothesis

is that determining genetic loci associated with multiple primary pain conditions will provide stronger evidence for their involvement in pain processes.

2.2. Methods

2.2.1. Aims

The main aim of this systematic review is to query the available literature and identify variants potentially implicated in pain vulnerability in complex conditions.

The first aim was to evaluate the extent of overlap between genetic loci that are genome-wide significantly associated with primary pain conditions, by reviewing data from GWAS studies and performing correlation analysis, taking account of linkage disequilibrium between variants at the loci.

The second aim was to investigate whether primary pain associated variants are also in linkage disequilibrium with GWAS hits from endometriosis, osteoarthritis, RA, IBD.

The third aim was to review the biology of these variants of interest in the context of chronic pain mechanisms.

2.2.2. Eligibility criteria

This review is limited to genome-wide association studies due to the large number of participants, stringent quality control and high coverage of the human genome. There were no participant restrictions based on age, sex, or ancestry.

Studies were included if they sought to determine associations between a primary pain condition and genetic variants. Primary pain conditions for which there are available genome-wide association studies are shown in Table 1.

Table 1. List of primary pain conditions for which genome-wide association studies have been performed and are available in GWAS Catalog[67].

Trait	No. of genome-wide association studies	No. of independent associations	Study size average (range)
Chronic back pain	2	7	313,145 (158,025 - 468,265)
Fibromyalgia	1	0	503
Chronic widespread pain	2	0	133,327 (16,815 - 249,843)
Multisite chronic pain	3	50	258,433 (178,556 - 387,649)
Complex regional pain syndrome	1	0	230
Irritable bowel syndrome	5	4	163,907 (1,570 - 455,321)
Dysmenorrhea	2	5	8,812 (5,734 - 11,891)
Migraine	10	49	77,700 (285 - 375,752)
Cluster headaches	2	4	4,324 (458 - 8,191)
Temporomandibular joint disorder	3	3	13,703 (3,030 - 10,153)

2.2.3. Search strategy

The systematic review was registered and accepted for inclusion in PROSPERO (Lucinescu *et al*., 2021) in August 2021 (PROSPERO ID Number: CRD42021270393).

Electronic searches were conducted in the GWAS Catalog, which is a vast, curated database of peer reviewed GWAS, updated regularly and publicly available[67]. The GWAS Catalog search was performed four times between the 26th of July 2021 and 2nd of November 2021. The search terms were “chronic pain”, “pain”, “irritable bowel syndrome”, “headache” and “temporomandibular joint disorder”.

2.2.4. Selection process

GWAS Catalog performs weekly literature scans of studies that include a primary GWAS analysis, defined as array-based genotyping and analysis of 100,000+ pre-QC SNPs selected to tag variation across the genome and without regard to gene content. Exclusion criteria are studies published in a language other than English, candidate gene studies, studies using customized gene-based arrays without a clearly described GWAS methodology and studies that measure somatic variation (e.g., cancer). Every study is assessed for eligibility by independent curators. Literature scans were formerly performed manually in Pubmed, and since 2018, they have been perfomed using a machine learning classifier[67].

This study does not include any new GWAS data. Studies need to have publicly available summary statistics to be included. The GWAS Catalog uses traits to categorise similar studies. Traits are ontology terms used to classify the outcome, such as disease, phenotype, measurement, or drug response. Studies represent entries in the GWAS Catalog. Multiple traits may be registered for the same publication.

Searches were performed for each term. Then, duplicate traits/studies were removed. Next, the studies were screened using a two-step process. Firstly, the studies were screened by Title/Abstract for eligibility. Secondly, the remainder of studies were retrieved and subjected to a full-text screen. A consensus check was completed by the rest of the reviewers (my supervisors) to verify whether the identified traits/studies were eligible.

2.2.5. Quality assessment of studies

Quality assessment of the studies was performed once at the time of curation in the database and once after screening, using a modified STrengthening the REporting of Genetic Association studies (STREGA) framework (Supplementary Tables 1 and 2). The STREGA initiative builds on the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement[68]. It adds genetic association study specific checklist items to help reporting in these studies and in assessing the strength of evidence.

2.2.6. Data items and collection process

Relevant information for each trait was extracted directly from the GWAS Catalog database. Extracted information included: associated SNPs, chromosomal location, risk allele, p-value, odds ratio, beta, confidence intervals, mapped gene, reported trait and the trait registered in the GWAS Catalog[67].

Information about the studies was extracted and consisted of publication date, journal, title, reported trait, trait(s) assigned by GWAS Catalog, discovery sample description, replication sample description, discovery sample number and ancestry, replication sample number and ancestry, and association count[67].

2.2.7. Linkage disequilibrium between variants

We measured correlation between biallelic variants from the GWAS located in genomic proximity to one another by calculating r^2 within the 1000 genomes reference panel using the LDlink website online tool LDpair[69,70].

Reference populations were: African (African/African American/Afro-Caribbean), Ad Mixed Americans (Latin/Hispanic), East Asian (East Asian) and European (European and Norfolk Islanders).

2.2.8. Data synthesis

The list of variants was subsetted based on ethnicity of participants. Then, each new table was sorted based on chromosomal location. For each pair of variants, r^2 and D' were calculated using the LDpair option on LDlink website using the correct reference population (see above)[70]. D' is a measure of linkage disequilibrium whereas r^2 is a measure of correlation between SNPs that takes into account allele frequency. We can only talk about correlation if the SNPs are on the same chromosome, therefore variants on different chromosomes were excluded. Non-biallelic variants or variants not on human genome build 38 were excluded. A threshold of 0.8 for r^2 was considered moderate to high evidence of correlation, according to literature, whereas 1.0 is considered complete correlation (the SNPs are the same).

For the list of selected variants, Phenoscanner was queried for reported associations ($P < 5 \times 10^{-8}$) with other traits and conditions (http://www.phenoscanner.medschl.cam.ac.uk/; July 2021). PhenoScanner is a curated database holding publicly available results from large-scale GWAS[71,72]. Parameters were set to an r^2 of 0.5, European ancestry as reference population and the p-value threshold at 5×10^{-8}, as customary for GWAS studies. The r^2 threshold selected indicates modest evidence of correlation.

The list of variants was then annotated in Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) using SNP2GENE function[73]. The SNP2GENE function takes GWAS summary statistics as an input and provides functional annotation for all SNPs in genomic areas identified by lead SNPs. FUMA accepts as input a comma-separated file of GWAS summary statistics with the rsID, p-value, chromosomal position, and effect allele. In addition, beta and odds ratio

values were provided where known. A p-value threshold of lead SNPs was set to 5 x 10^{-8} and the reference panel population was set to 1000G Phase3 EUR. Only positional gene mapping was selected. Next, GENE2FUNC function was selected with the list of suggested gene mappings provided by SNP2GENE. GENE2FUNC function provides gene specific annotation by producing a heatmap, calculating differentially expressed gene sets in different tissue types and providing biological context for the genes. The rest of the analyses were performed using R Statistical Software (v4.1.2; R Core Team 2021)[74]. The R package phenoscanner was used to query the PhenoScanner database[71,72].

Significant variant gene associations were downloaded from Genotype-Tissue Expression (GTEx) Portal Release V8 on 1 February 2022[75]. Genes regulated by the variants of interest or variants in LD with them were extracted, alongside p-value, tissue, and gene information. Then, the Human Pain Genetics Database (HPGD) and two endometrium eQTL studies were searched with the list of variants of interest[76–78].

2.3. Results

2.3.1. Study selection

The GWAS Catalog was queried for primary chronic pain conditions. Six hundred eighty-seven studies were returned, and 151 duplicate studies were removed. After screening the Title/Abstract, 410 studies not related to chronic pain were excluded because they did not meet inclusion criteria. The remainder of 126 studies were retrieved and subjected to a full-text screen. Seventy-seven studies were excluded for not investigating primary chronic pain traits. Finally, a total of 49 studies met the

eligibility criteria (Figure 1). For a full list of included studies and quality assessment, see Supplementary Tables 2 and 3.

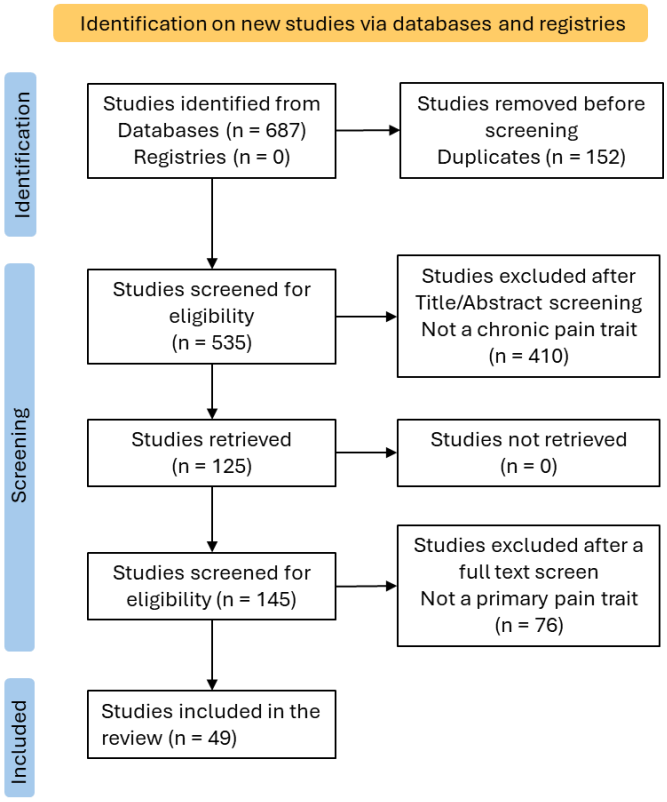


Figure 1. PRISMA flow-chart of the systematic review.

2.3.2. Study characteristics

In terms of ancestry of the participants, 39 (79.59%) of the studies had European participants, 11 (22.44%) had East Asian participants, 3 (6.12%) had African American, African, or Afro-Caribbean participants and 3 (6.12%) had Hispanic/Latin American ancestry.

2.3.3. Evidence of correlation between primary pain conditions

GWAS in primary pain conditions yielded either no significant associations ($p < 5 \times 10^{-8}$), in the case of fibromyalgia, chronic widespread pain or complex regional pain syndrome, or many independent associations, in the case of multisite chronic pain or migraine (Table 1). Associations extracted from the GWAS Catalog were first filtered

to remove variants with a p-value greater than than 5×10^{-8}. Table 2 shows the list of variants that represent the same loci (through LD > 0.8) associated with more than one primary pain condition.

2.3.4. Evidence of genetic correlation between primary pain conditions and other traits

Next, other traits associated with the final list of variants were extracted from Phenoscanner database[71,72]. The resulting traits were categorised into vascular, haematological, pharmacological, respiratory, immunological, anthropometric, and other (see Supplementary Table 4). The most frequently reported associations between a trait and the variants of interest were pharmacological (most common being “Medication for pain relief, constipation, heartburn: none of the above” and “Medication for pain relief, constipation, heartburn: paracetamol”) (Figure 2).

Table 2. List of variants in linkage disequilibrium associated with different primary pain traits ($r^2 > 0.8$).

Variant	Position	Mapped gene	Trait 1	Trait 2
rs12134493	1:115134562	LINC01765	Migraine	Headache
rs7523086, rs7544256	1:115280766	NGF-AS1	Dysmenorrhea	Migraine
rs10166942, rs2362290	2:233916086	TRPM8, MSL3P1	Migraine	Headache
rs9349379	6:12903725	PHACTR1	Migraine	Headache
rs7775721, rs11153082	6:96609103, 6:96611790	FHL5	Migraine without aura	Cluster headache
rs6478241	9:116490350	ASTN2	Migraine without aura	Multisite chronic pain
rs11172113	12:57133500	LRP1	Migraine	Headache

Variant rs7523086 intronic to *NGF* gene correlates with a variant (rs12030576) associated with endometriosis in the largest meta-analysis of endometriosis GWAS to date (r^2 = 0.978)[40].

FUMA was queried with the list of variants from Table 2 to annotate the variants and to map them to genes[73]. Then, Phenoscanner database was queried with the list of genes provided by FUMA (Table 3)[71,72]. Results implicate *NGF*, *LRP1* and *ASTN2* in rheumatoid arthritis, and *ASTN2* in ulcerative colitis and osteoarthritis. According to the heatmap of tissue-specific gene expression from FUMA, *ASTN2* and *LRP1* are highly expressed in brain tissues (Figure 3). Although lowly expressed in brain tissues and spinal cord, *NGF* is expressed in various gynaecological tissues, including the uterus, fallopian tubes, and cervix.

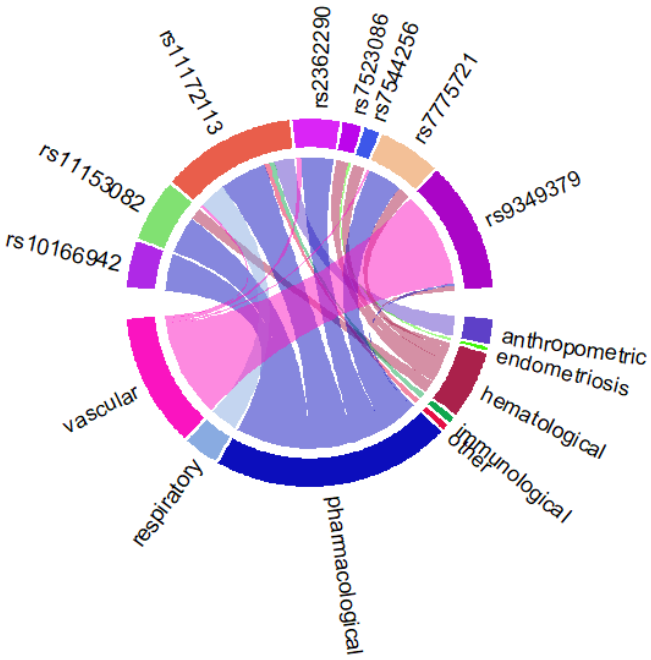


Figure 2. Chord plot of variants of interest (top, rsids) and the traits they are associated with according to Phenoscanner[71,72].

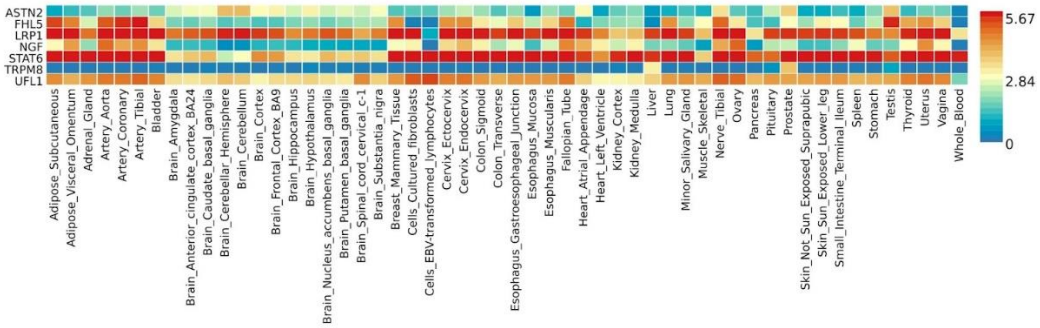


Figure 3. Heatmap of genes of interest in 53 tissue types from GTEx version 8. Figure was produced by FUMA[73].

Table 3. Mapped genes for the variants of interest with the FUMA tool with SNP2GENE option[73].

Variant ID	Location	Mapped gene
rs12134493	1:115677946	n/a
rs7523086	1:115280766	NGF
rs10166942	2:233916086	n/a
rs9349379	6:12903725	n/a
rs7775721	6:96609103	n/a
rs6478241	9:116490350	ASTN2
rs7544256	1:115280766	n/a
rs2362290	2:233916086	TRPM8
rs11153082	6:96611790	UFL1, FHL5
rs11172113	12:57133500	LRP1, STAT6

2.3.5. Expression quantitative trait loci analysis

The variants of interest regulate the expression of many genes in different tissues, however no relevant tissue eQTLs were highlighted by the GTEx analysis[75]. Dorsal root ganglia are involved in ascending transmission of pain signals and descending modulation of pain, therefore the Human Pain Genes database of DRG eQTLs was investigated. Similarly, since one variant of interest is associated with endometriosis,

two endometrium eQTL databases were investigated. No statistically significant dorsal root ganglia or endometrium eQTLs were found[76–78].

2.4. Discussion

2.4.1. Key findings

In this systematic review of the published literature, GWAS were performed for a range of primary pain conditions, with different success rates in identifying associated loci. Migraine and multisite pain are well represented traits in terms of case numbers, with many associated genetic variants, whereas fibromyalgia GWAS was underpowered and did not identify any genome-wide significant associations. Amongst the correlated variants between different primary pain conditions, many were between migraine and chronic headaches, one between multisite pain and migraine and one between dysmenorrhea and migraine. Other traits associated with the correlated variants included many anthropometric, pharmacological, and immunological traits. One variant intronic to *NGF* was also associated with endometriosis from a recent endometriosis meta-analysis which at the time of the review was not part of the GWAS Catalog[40]. FUMA mapped these variants to *NGF*, *LRP1* and *ASTN2*. No relevant eQTLs could be identified in brain, endometrium, or dorsal root ganglia; however there is evidence of these genes being expressed in brain and reproductive tissues.

In this systematic review, these genes were shown to influence multiple pain traits and may be involved in a common pain vulnerability framework for chronic pain conditions: *LRP1* (associated with migraine, headache, and RA), *NGF* (associated with dysmenorrhea, migraine, RA and endometriosis) and *ASTN2* (associated with migraine, multisite chronic pain, osteoarthritis, RA and ulcerative colitis).

NGF has been previously hypothesised to contribute to pelvic pain in endometriosis by sensitisation of afferent pathways in the peritoneal cavity[59]. Moreover, *NGF* expression is upregulated in cerebrospinal fluid of migraine patients[79]. There is evidence of central sensitisation being an important pain mechanism in migraine and endometriosis[8,80,81]. Therefore, *NGF* could be involved in both central and peripheral sensitisation mechanisms. Tissue specific gene expression in GTEx showed *NGF* is expressed in gynaecological organs, consistent with a peripheral sensitisation mechanism in chronic pain. Expression in brain tissues appeared low.

LRP1 is a cell surface receptor that is involved in the proliferation of vascular smooth muscle cells and plays a role in indirectly decreasing neurodegeneration and neuropathic pain by attenuation of the neuroinflammatory environment. A study by Alban *et al*., (2008) has shown *LRP1* activation can indirectly decrease neuropathic pain by attenuation of the inflammatory environment through inhibition of MAPK c-Jun N-terminal kinase and *p38*-pathway[82]. By inhibiting these pathways, *LRP1* represses expression of TNF-α, IL-1β, and IL-6. Many neuropathies are associated with TNF-α and IL-6 activation. Interestingly, *MAPK* pathway activation is observed in endometriosis and is thought to be involved in every step of the endometrioma formation: migration, implantation, growth, and proliferation[83]. Moreover, *LRP1* polarizes microglial cells and macrophages from a proinflammatory phenotype (M1) to an anti-inflammatory phenotype (M2), attenuating the neuroinflammatory environment[84]. Its role in migraine and chronic headaches remains unknown, but *LRP1* appears to be highly expressed in brain tissues and spinal cord, according to GTEx (v8)[75].

A*STN2* plays a role in glial-guided migration that appears important for development of the laminar architecture of cortical regions in the brain. A study by Inoue *et al*.,

(2021) implicated two SNPs within *ASNT2* in opioid analgesia. Its role in chronic pain is presently unknown[85]. A*STN2* appeared to be expressed moderately in brain tissues, according to GTEx (v8)[75].

2.4.2. Limitations

There are limitations of this systematic review. Firstly, it is unlikely that GWAS can uncover all genetic determinants of a disease. The inheritance of complex diseases is thought to be only in part the result of common genetic variants. Rare variants also play an important role. The genetic variant burden model illustrates one of the proposed mechanisms (see Chapter I)[86]. Moreover, discovery of common genetic variants associated with chronic pain requires very large numbers of cases, which so far have not been achieved. Some chronic pain GWAS were underpowered (*e.g.*, fibromyalgia). Secondly, there are differences in diagnostic criteria for primary pain conditions between different studies, which can impact the study selection process and consequent results.

2.4.3. Conclusions and Future directions

This systematic review of the current literature on primary pain variants involvement in secondary chronic pain points to *NGF*, *LRP1* and *ASTN2* as potential chronic pain vulnerability genes. Here, I show they have been associated with both primary and secondary pain conditions and are expressed in relevant tissues for chronic pain. Future research should focus on understanding how these genes are involved in chronic pain pathways and how they give rise to comorbid chronic pain conditions.

Chapter III: Genetic analysis of fibromyalgia in UK Biobank

3.1. Rationale

The previous chapter identified that the largest published GWAS of fibromyalgia to date was underpowered (<500 cases) and did not identify any genome-wide significant hits ($p < 5 \times 10^{-8}$)[87]. UK Biobank and FinnGen are large databases of 500,000 participants, each providing genomic and diagnostic data[88,89]. In order to boost power of the detection of genetic associations with fibromyalgia, I conducted a large, sex stratified GWAS of ~ 5,000 fibromyalgia cases from UK Biobank and a meta-analysis with a GWAS of fibromyalgia of ~3,000 cases from FinnGen. Furthermore, I investigated the heterogeneity of different fibromyalgia case definitions in UK Biobank and compared different GWAS methods to select the most appropriate one.

3.2. Methods

3.2.1. The UK Biobank

The UK Biobank is a large-scale biomedical database and research resource containing genetic, lifestyle and health information from half a million UK participants[88]. Cases were identified using self-reported data (collected at recruitment or in an online follow-up questionnaire), hospital diagnosis data (by data-linkage to hospital records with ICD10/ICD9 codes) and primary care records (by linkage to general practice data with read codes v2 and v3). At the Biobank recruitment visit (2006-2010), and in subsequent UK Biobank clinic follow-ups in 2012-2013 or 2014-2019, participants were asked via touch screen questionnaire about ‘serious medical conditions previously diagnosed by a doctor’. If participants reported a medical condition, they are asked during a verbal interview with a nurse to

state the condition(s) (Data field 20002). These participants were referred to as unprompted self-reports. Participants could also be diagnosed with a condition during a hospital admission and coded in the primary or secondary ICD-10 diagnosis or primary or secondary ICD-9 diagnosis (Data fields 41202, 41203, 41204, 41205). These participants were referred to as hospital ICD9/ICD10 cases. Participants also completed follow-up questionnaires regarding their pain (Category 154). Ethnicity was reported under Field 21000 (White British: 1001).

3.2.2. Participants

Fibromyalgia cases were defined by combining data from hospital in-patient diagnosis of fibromyalgia either recorded with ICD10 (M797) or the equivalent ICD9 codes (72910-72919), from self-reported answer in the baseline questionnaire (Code 1542), from self-reported previous diagnosis of fibromyalgia (Field 120009, “Ever had fibromyalgia syndrome”) or from primary care data that is collected from events that general practice staff record to support clinical care (Table 4).

Most fibromyalgia cases from the ICD cohort come from codes M79.79 (Fibromyalgia at unspecified site: 421 cases) and M79.70 (Fibromyalgia at multiple sites: 83). Most fibromyalgia cases from primary care were driven by codes: X75rx (1666 cases), N248. (381), N2401 (278), N2139 (177) and N2121 (126).

3.2.3. UK Biobank GWAS of fibromyalgia

Due to the low numbers of other ethnic groups in UK Biobank (484 non-White or other White, 5433 White British), only White British cases were included. Sex-stratified (females-only and males-only) GWAS was conducted in BOLT-LMM and REGENIE. Sex-combined GWAS was conducted in BOLT-LMM, REGENIE and SAIGE to compare results.

Table 4. Fibromyalgia case and control definitions in UK Biobank for genome-wide association study.

Filter description	N cases	N controls	Code	Code name
Hospital ICD-10	518	487859	M797	Fibromyalgia
Hospital ICD-9	532	487845	72910:72919	Myalgia and myositis, unspecified
Unprompted self-reported	1389	486988	1542	Fibromyalgia
DHQ self-report Pain	3922	484455	120009	Ever had fibromyalgia syndrome (Yes)
Primary Care data	5917	482460	N240., N2401, N2405, N2406, N240z, X708C, X75rx, .M4A9, N239., N248., N2480, XE1IQ, .M4B., .M4BZ, N2121, N240z	Fibromyalgia or fibrositis
Total all ancestry	5917	482460		
White British ancestry	5433	409782		
Sex-combined GWAS	5433	409782		
Females-only GWAS	4512	215744		
Males-only GWAS	921	228148		

Genotype quality control (QC) and imputation was performed in UK Biobank centrally. The genotype QC and imputation is described in detail in Bycroft *et al.* They provide a list of samples failing QC due to missingness, sex mismatch and/or high heterozygosity which were excluded from further analysis[90]. Participants who retracted their consent were also excluded. Genotype QC briefly included keeping only those variants with an imputation quality score of more than 0.3, minor allele frequency (MAF) >= 0.01, genotypes without significant deviation from Hardy-Weinberg Equilibrium test ($p > 10^{-6}$), biallelic and non-ambiguous SNPs. Additional genotype QC for (1) SAIGE: a minor allele count cut-off of at least 20 were kept; (2) REGENIE: the genotype files were processed using PLINK2 to filter out SNPs with minor allele frequency (MAF) below 1%, genotype missingness above 10% and

Hardy-Weinberg equilibrium p-value exceeding 10^{-15}, and samples with more than 10% missingness[91–93].

BOLT-LMM has been historically successfully applied to thousands of phenotypes in UK Biobank, yielding many genome-wide significant associations. It uses a Bayesian linear mixed effects model to fit the data. It was shown to be reliable and fast when applied to large biobanks. However, linear mixed models can result in inflated Type I error rates due to unbalanced case-control ratios, like in the case of most binary traits in UK Biobank[94]. BOLT-LMM was performed using array type as covariate. BOLT-LMM computes a genetic relationship matrix at the time of the computation, therefore principal components were not included as covariates[94]. The command was run using *--lmmForceNonInf* option, which computes an association statistic for both infinitesimal and non-infinitesimal models, however, for the rest of the analysis, only the non-infinitesimal model was considered[94].

SAIGE uses a logistic regression mixed model with saddle-point approximation, which is meant to minimise false positives and improve power of detection of true associations, even at small allele frequencies. SAIGE for binary traits fits a null logistic mixed model, calculates a genetic relationship matrix (GRM), and performs single-variant association analyses with saddle-point approximation. Step one of SAIGE consists of deriving the genetic relationship matrix and fitting the model, using a pruned list of SNPs from PLINK2 software[92]. Covariates included were array type and the first 10 principal components. Options used were: *LOCO = FALSE* and *tauInit = 1,0*. Step two involves performing single-variant association tests with saddle-point approximation, which accounts for case-control imbalances, such as the ones occurring in UK Biobank.

REGENIE comprises two steps: step one uses a whole genome regression model in blocks of SNPs to reduce the number of dimensions; then in step two it tests a larger set of markers (from imputation) for association with the phenotype based on predictions from step one. It is the faster than SAIGE and has similar effect size estimates when applied to binary traits[95]. Covariates included array type and the first ten principal components. Step one was ran with binary trait option (--bt), and genotype block size 1000 (--bsize 1000). Step two was run with binary trait option (--bt), genotype block size 400 (--bsize 400) and approximate Firth likelihood ratio test for p-values less than 0.01 (--firth --approx --pThresh 0.01), which is faster than exact Firth correction. The Firth correction removes much of the bias that arises when using maximum likelihood estimates, and compared to saddle-point approximation, it controls for inflated effect size estimates[95].

Sex-combined, female specific and male specific signals were identified by extracting lead SNPs utilising FUMA ($p < 5 \times 10^{-7}$), then comparing p-values and effect sizes between females, males, and sex combined populations. If a lead SNP was nominally significant ($p < 10^{-5}$) in females but not in males, it was considered a female-specific signal, and vice versa.

3.2.4. Sensitivity analysis

SNP-based heritability is the proportion of variation in the trait that can be explained by additive effects of commonly occurring genetic variants. There is limited data on SNP-heritability of chronic pain conditions in UK Biobank, available from the Neale Lab Heritability browser[96]. From their data, there are differences in SNP-heritability between different data sources: ICD-10, ICD-9 and self-reported. Low SNP-heritability in self-reported data could be caused by a lack of proper case definition. A previous study has shown self-reported data in UK Biobank is correlated with

hospital data and combining them reliably improves power to detect associations[97]. To address the potential discrepancies between hospital and self-reported data, sensitivity analyses were conducted involving separate GWAS of the different case definitions.

The goal of sensitivity analyses was to ascertain whether self-reported cases, and primary care cases are congruent with hospital-episode identified cases of fibromyalgia in terms of genetical correlation and effect size of top signals. Briefly, all cases with a clinical diagnosis of fibromyalgia, either through primary care or hospital in patient diagnosis, were allocated to one case group; self-reported cases, either at recruitment or in follow-up questionnaires, were assigned to a second group, with the controls (the UK Biobank cohort with no record of fibromyalgia) split evenly and randomly between the two case groups. There were 3131 clinically confirmed fibromyalgia cases and 2627 self-reports. Then, the two summary statistics were used to compute pair-wise genetic correlation using Linkage Disequilibrium Score Regression (LDSC) (see below). In the next sensitivity analysis, all cases with a primary care level diagnosis (N = 2788) were assigned to one group and hospital diagnoses were assigned into a second group (N = 532). Controls were filtered out for different sources of fibromyalgia diagnosis and split evenly and randomly into two cohorts.

3.2.5. FinnGen

FinnGen is a large public-private collaboration which combines genomic data with digital healthcare records of 500,000 participants[89]. Samples include previous disease-specific research cohorts (200,000) and prospective participant samples (300,000), from biobanks across Finland. The collaboration brought together samples from six regional hospitals, private healthcare and blood donation biobanks.

3.2.6. Participants

The FinnGen population has been described elsewhere[89]. The definitions of FinnGen disease end points and their respective controls for each release are available at https://www.finngen.fi/en/researchers/clinical-endpoints, and FinnGen end points can also be browsed at https://r5.risteys.finngen.fi/.

Fibromyalgia cases were defined using HILMO registry (in-patient and outpatient primary and secondary diagnoses: ICD10) (M79.7) and Cause of death registry (immediate, underlying and contributing causes of death on the death certificate with ICD10 codes). In total, there are 3284 Finnish participants with diagnosed fibromyalgia. After genotype quality control, there were 3166 cases, of which 2956 females and 210 males.

Quality control included filtering individuals with genetically inferred sex not matching the reported sex in registries, high genotype missingness (>5%) and excess heterozygosity (±4 standard deviations) and filtering variants with high missingness (>2%), low Hardy–Weinberg equilibrium ($P < 1 \times 10^{-6}$), minor allele count < 3, imputation INFO scores of <0.6 and MAF values of <0.01.

3.2.7. FinnGen GWAS of fibromyalgia

Sex-combined GWAS of fibromyalgia was performed by FinnGen using the whole genome regression model method REGENIE (version 2.2.4), using sex, age, genotyping batch and ten principal components as covariates[95]. Summary statistics of fibromyalgia GWAS are publicly available from FinnGen Release 9.

3.2.8. Meta-analysis of FinnGen and UK Biobank fibromyalgia

The ICD-based diagnosis of fibromyalgia sex-combined GWAS results from UK Biobank and FinnGen were included in a meta-analysis (N_{cases} = 3,698; $N_{controls}$ =

699,009). Meta-analysis of 6,109,036 markers was carried out using the inverse variance weighting fixed-effects method implemented in METAL[98]. First, variants were filtered at the study level for MAF > 1% and MAF < 99%, and non-ambiguous, biallelic SNPs. METAL was run with default options[98]. The results were genomic controlled. A filter was introduced to remove SNPs which were present in less than half the total sample size to ensure the SNPs are present in both datasets. A meta-analysis p-value of $< 5 \times 10^{-8}$ was selected as the significance threshold for association.

3.2.9. Genetic correlation

LD score regression was performed using LDSC v1.0.1 (https://github.com/bulik/ldsc)[99]. The LDSC command-line tool was run using precomputed European LD scores available from Broad Institute (https://data.broadinstitute.org/alkesgroup/LDSCORE/eur_w_ld_chr.tar.bz2), with calculations limited to SNPs from HapMap3 (https://data.broadinstitute.org/alkesgroup/LDSCORE/w_hm3.snplist.bz2)[100]. The tool allows computing LD scores using GWAS summary statistics, SNP-heritability estimates, and genetic correlation between multiple traits. Summary statistics were prepared for analysis using the munge_sumstats.py function of LDSC.

3.3. Results

3.3.1. Cohort characterisation

Fibromyalgia in UK Biobank is characterised by a middle aged, female preponderance, of similar BMI and educational background and who mostly self-report as having at least a fair overall health (Table 5). The ICD only cohort and the clinical cohort have reported significantly worse health ratings than the primary care

only cohort and self-reported respectively, using the Pearson's X^2 test with one degree of freedom (ICD - Primary care: $X^2 = 120.76$, p-value = 2.2×10^{-16}; Clinical - Self-reported: $X^2 = 51.89$, p-value = 5.88×10^{-13}). The clinical cohort also reported differences in educational attainment compared to the self-reported cohort, using Pearson's X^2 test with seven degrees of freedom ($X^2 = 294.65$, p-value = 2.20×10^{-16}). No statistically significant differences were found between ICD-only cohort versus primary care-only cohort ($X^2 = 5.2$, p-value = 0.64). Pearson's X^2 test with one degree of freedom identified a statistically significant sex difference between ICD-only and primary care-only cohorts, however the percentage of females to males is very similar (82.7% versus 82.2%). Therefore, this does not suggest a clinically significant finding. Similarly, when conducting Welch's two-sample t tests for age and BMI, there appear to be statistically significant differences, however examining the means does not prompt to a clinically meaningful observation that can be drawn from these differences (ICD - Primary care: $t_{Age} = -2.89$, $p_{Age} = 4.00 \times 10^{-3}$, $t_{BMI} = 5.31$, $p_{BMI} = 1.54 \times 10^{-7}$; Clinical - Self-reported: $t_{Age} = 3.59$, $p_{Age} = 3.32 \times 10^{-4}$, $t_{BMI} = 4.38$, $p_{BMI} = 1.22 \times 10^{-5}$)(Table 5).

3.3.2. Genome-wide association analysis in UK Biobank

To detect sex specific genetic differences in fibromyalgia, GWAS were run separately for males (N = 921), females (N = 4512) and sex-combined (N = 5433) in UK Biobank including cases ascertained from clinic and self-reported resources to maximise power of detection (Figure 4). Lambda values for genomic control were: $\lambda = 1.048$ (sex combined), $\lambda = 1.047$ (females) and $\lambda = 1.00$ (males).

Table 5. Sociodemographic descriptives of fibromyalgia cases in UK Biobank by case ascertainment category.

Item	ICD	Primary care	Clinical (ICD + Primary care)	Self-reported
Total	532	2788	3131	2627
Demographics				
Females	440 (82.7%)	2291 (82.2%)	2587 (82.6%)	2150 (81.8%)
Males	67	497	544	477
Age	55	56	57	55
Ethnicity (White British)	442	2479	2857	2503
BMI	30.6	28.93	28.42	29.09
Health Rating (%)				
Fair and up	49.81	75.90	73.52	81.80
Poor	44.36	23.06	25.42	17.62
N/a	5.83	1.04	1.05	0.57
Education Level (%)				
A level (and above)	29.32	30.42	32.06	12.37
GCSE	24.43	22.63	24.30	5.08
Other (vocational)	18.05	19.3	20.25	69.81
None	22.37	25.79	26.76	11.27

In females, a total of 34 lead SNPs were identified at a more stringent nominal threshold of association ($p < 5 \times 10^{-6}$) and no genome-wide significant hits were identified. In males, one genome-wide significant SNP was discovered on chromosome 17 (rs75888727, $p = 3.9 \times 10^{-8}$). A total of 46 lead SNPS were identified for males ($p < 5 \times 10^{-6}$). In sex combined GWAS, one variant approached significance on chromosome 2 (rs146396094, $p = 8.9 \times 10^{-8}$) and 27 lead SNPs were identified ($p < 5 \times 10^{-6}$). Comparing the effect sizes for the top lead SNPs ($p < 5 \times 10^{-7}$) between males, females and sex-combined, revealed similar effect sizes and consistently small p-values ($p < 5 \times 10^{-5}$) between female and sex combined top associations (Table 6). Top male signals were not present in either female GWAS or

sex combined. This suggests distinct male and female genetic drivers of disease. The lead SNPs were located in non-coding regions, either intronic or intergenic, according to annotations performed in FUMA. Then lead SNPs were mapped positionally to genes by FUMA. The strongest signal for females and sex combined data sets was located in the *COMMD1* locus, and for males, in the *ARSG* locus. The top genes associated with fibromyalgia in females were: *COMMD1*, *RIBC2*, *INO80C* and *SLC39A11*. For males, the top genes were *ARSG*, *ANKRD20A9P*, *EPHA1-AS1* and *LOC105377114*.

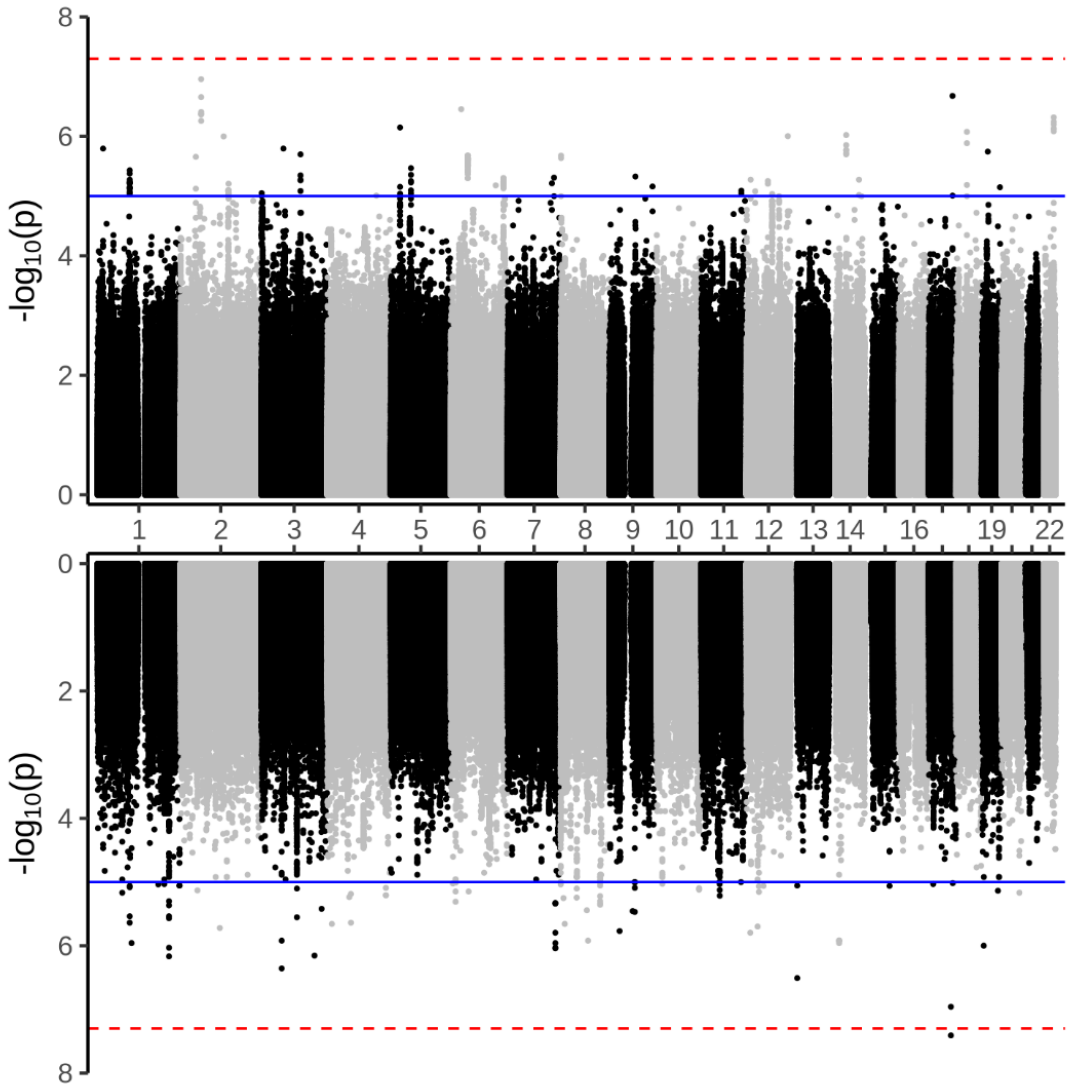


Figure 4. Sex-stratified Miami Plot of fibromyalgia in UK Biobank. Female only GWAS results given as a Manhattan plot on the top and male only GWAS results given as a Manhattan plot on the bottom. Each dot on the plot is a SNP and significance of association is given on Y-axis. Red-dashed line represents genome-wide significance threshold ($p<5x10^{-8}$), and the blue line represents a relaxed threshold for nominal significance ($p<5x10^{-5}$).

Table 6. Top lead SNPs ($p < 10^{-7}$) of female-only, male-only, and sex-combined genome-wide association studies of fibromyalgia in UK Biobank.

					Female (N = 4314)			Male (N = 921)			Sex Combined (N = 5314)			
SNP	CHR	BP	EA	EAF	P	BETA	SE	P	BETA	SE	P	BETA	SE	Gene
rs11659874	18	33069644	T	0.75	8.40E-07	0.120	0.024	2.40E-01	0.080	0.068	5.50E-07	0.061	0.012	*INO80C*
rs7209250	17	71029699	T	0.84	2.10E-07	-0.150	0.029	3.00E-01	-0.083	0.081	2.80E-07	-0.160	0.030	*SLC39A11*
rs75888727	17	66394252	A	0.99	1.87E-03	0.012	0.098	3.90E-08	-1.486	0.270	1.11E-03	-0.247	0.105	*ARSG*
rs146160292	13	19447663	T	0.97	5.20E-02	-0.149	0.076	3.10E-07	-1.097	0.214	1.10E-04	-0.319	0.083	*ANKRD20A9P*
rs11763230	7	143108841	C	0.79	9.00E-01	0.003	0.025	9.10E-07	0.351	0.071	2.80E-02	0.061	0.028	*EPHA1-AS1*
rs147685768	3	61378601	A	0.98	7.80E-01	-0.021	0.074	4.40E-07	-1.045	0.207	1.80E-02	-0.189	0.080	*LOC105377111*
rs191424008	1	214953032	G	0.97	9.70E-01	-0.002	0.060	6.80E-07	-0.832	0.168	3.70E-02	-0.135	0.065	intergenic
rs7289955	22	45812562	T	0.67	4.80E-07	0.113	0.022	8.10E-01	0.015	0.063	4.60E-06	0.111	0.024	*RIBC2*
rs2185301	14	51441277	T	0.71	9.50E-07	-0.113	0.023	5.20E-01	-0.042	0.065	2.30E-06	-0.118	0.025	intergenic
rs118050321	12	125064554	C	0.93	9.90E-07	-0.201	0.041	7.60E-01	0.035	0.115	2.10E-05	-0.189	0.044	intergenic
rs55677935	2	62228013	C	0.98	1.10E-07	-0.438	0.083	6.90E-01	-0.095	0.234	3.40E-07	-0.457	0.090	*COMMD1*
rs146396094	2	62278808	A	0.98	2.20E-07	-0.422	0.081	2.40E-01	-0.273	0.230	8.90E-08	-0.219	0.041	*COMMD1*
rs76511846	3	4005132	G	0.90	1.90E-05	-0.150	0.035	na	na	na	9.10E-07	-0.090	0.020	intergenic

3.3.3. BOLT-LMM, REGENIE and SAIGE comparison

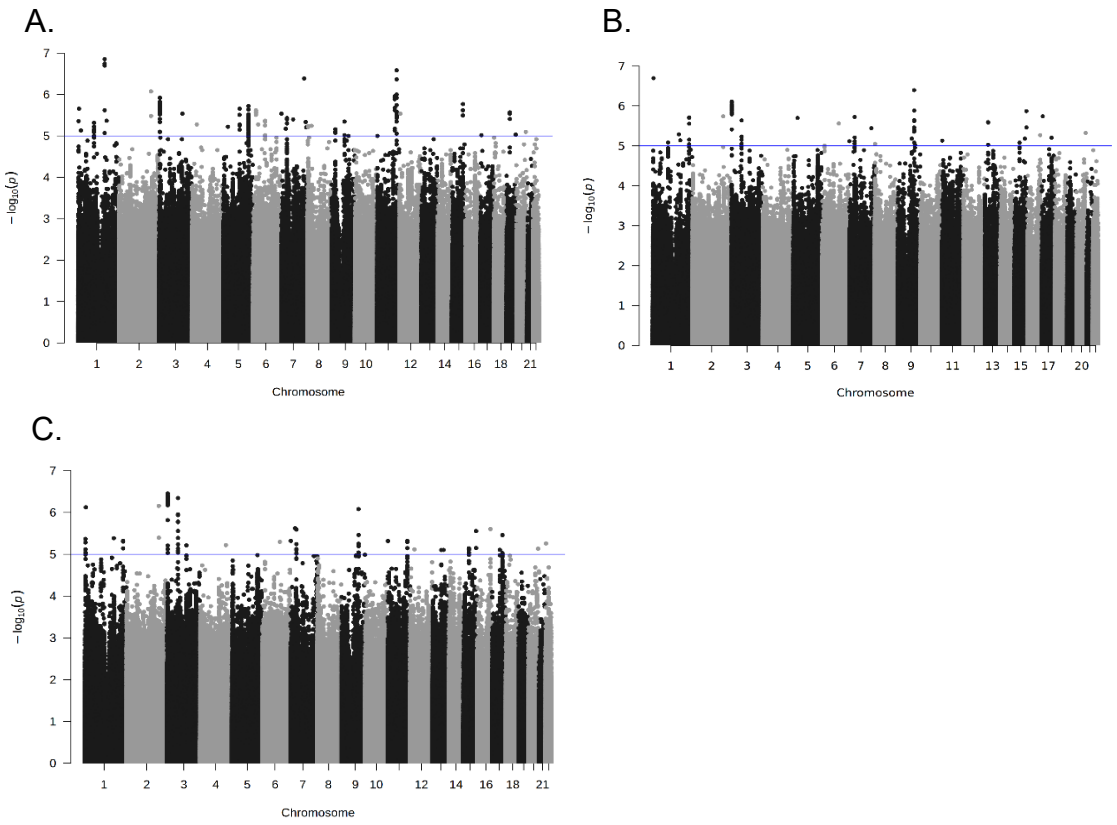


Figure 5. Manhattan plots of clinical diagnosis of fibromyalgia GWAS using different methods: BOLT-LMM (A), REGENIE (B) and SAIGE (C). Blue line suggests a nominal threshold of significance ($p<10^{-5}$).

Here, I have compared my results from BOLT-LMM with SAIGE and REGENIE which showed mostly similar results in terms of p-value and effect size of lead SNPs. The three methods were compared in terms of effect size, standard error, and p-value (Figure 5 and Table 7). Due to the computational speed of REGENIE and its previously reported lower inflated effect sizes, this method was utilised for the rest of the analysis.

Table 7. Independent lead SNPs ($p<10^{-7}$) from the assessment of three GWAS methods using clinical fibromyalgia UK Biobank cases. EA, effect allele. Missing values for several SNPs in SAIGE columns are due to them failing extra QC that SAIGE employs.

					BOLT-LMM			SAIGE			REGENIE		
NP	CHR	BP	EA	EAF	P	BETA	SE	P	BETA	SE	P	BETA	SE
rs148959249	3	3978283	A	0.055	1.50E-06	0.295	0.061	3.53E-07	0.286	0.056	8.31E-07	0.253	0.050
rs74702184	3	65627458	C	0.125	4.70E-05	0.168	0.041	4.52E-07	0.196	0.039	2.29E-06	0.174	0.036
rs72923554	2	193518421	T	0.042	8.40E-07	0.333	0.068	6.95E-07	0.323	0.065	1.81E-06	0.281	0.056
rs12131421	1	5472626	A	0.051	2.20E-06	-0.295	0.062	7.52E-07	-0.290	0.059	2.01E-07	-0.330	0.066
rs58228855	1	161671774	C	0.927	1.40E-07	-0.310	0.059	1.20E-05	-0.224	0.051	2.92E-05	-0.202	0.047
rs11216628	11	117794515	G	0.054	2.60E-07	0.354	0.069	1.99E-05	0.257	0.060	3.32E-05	0.233	0.054
rs143362105	17	71005035	A	0.990	2.80E-07	-0.742	0.144	1.39E-06	-0.696	0.144			
rs12672727	7	139354457	A	0.928	4.10E-07	0.275	0.054	1.11E-05	0.220	0.050	3.60E-06	0.247	0.055
rs10988763	9	102076659	A	0.051	1.00E-05	-0.127	0.029	8.31E-07	-0.131	0.027	4.01E-07	-0.136	0.027
rs977395	3	3968679	C	0.938	2.30E-06	-0.283	0.060	9.41E-02	-0.046	0.027	7.74E-07	-0.246	0.048

3.3.4. Sensitivity analyses by case ascertainment method

Sensitivity analyses were conducted to assess heterogeneity of the different case definitions of fibromyalgia in UK Biobank. Firstly, GWAS was conducted using clinically diagnosed cases of fibromyalgia only, and self-reports only. Top signals with the characteristic trail of SNPs underneath can be seen on chromosomes 1, 3 and 7 for clinical diagnosed fibromyalgia (Figure 6). Top signals from self-reported GWAS of fibromyalgia can be seen on chromosomes 8, 11 and 16 (Figure 6). QQ plots and genomic inflation factors were as expected: λ_{GC} = 1.047 (clinical diagnoses) and λ_{GC} = 1.029 (self-reports) (Figure 7). Lead SNPs appear not to be shared across the two GWAS (Table 8). Instead, the top signals for each GWAS appear to be unique to each cohort.

LDSC analysis showed the two data sets are 77.8% genetically correlated. SNP-heritability of clinically diagnosed fibromyalgia GWAS was 11.44%, whereas for self-reported cases it was 8.13%. A summary of sample characteristics between hospital-confirmed cases and self-reported cases is seen in Table 5. Both cohorts have a

similar age, female:male ratio, BMI, and ethnic background. Differences are observed for the self-reported overall health rating and educational attainment. Clinical cases appear to have worse self-reported health ratings.

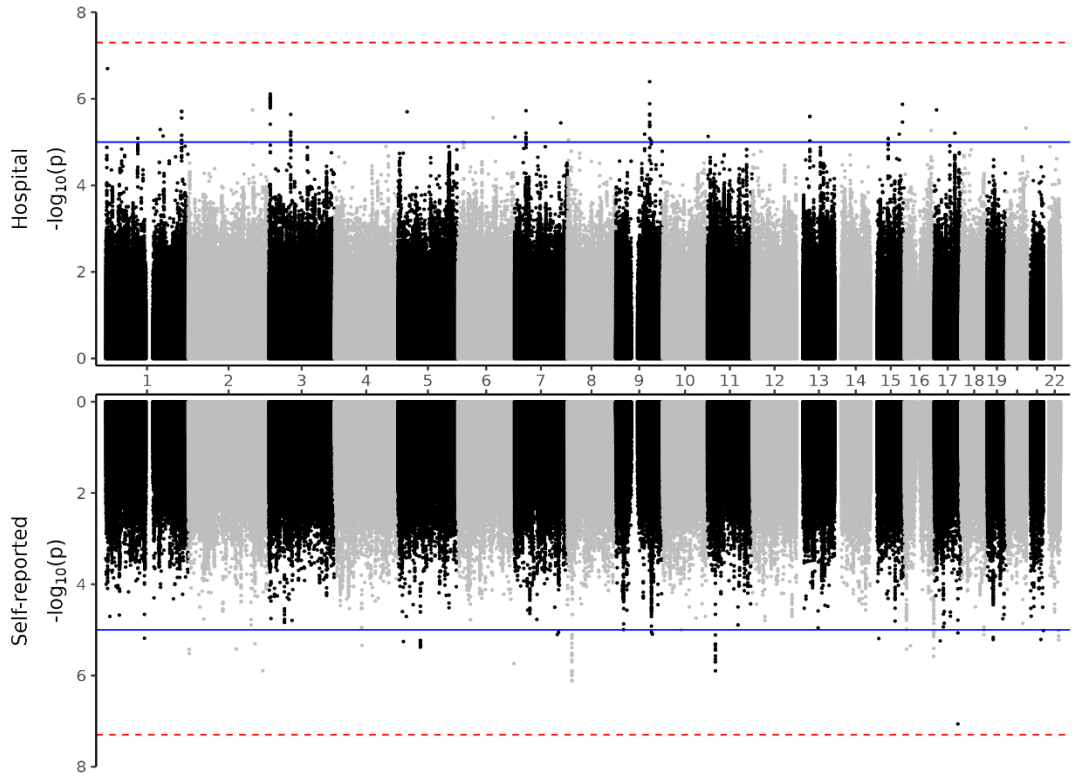


Figure 6. Miami plot of sensitivity analysis between clinical-confirmed fibromyalgia GWAS (top, labelled Hospital), and self-reported GWAS (bottom)

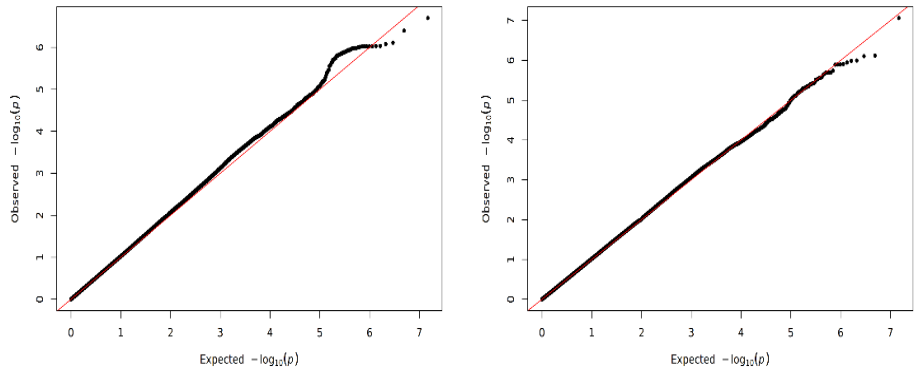


Figure 7. Q-Q plots of clinical-confirmed fibromyalgia GWAS (left) and self-reported fibromyalgia GWAS (right).

Table 8. Top 15 lead SNPs (p < 5 x 10-6) of clinical fibromyalgia GWAS and self-reported fibromyalgia GWAS. EA, effect allele.

				CLINICAL (HES+PRIMARY)			SELF-REPORTED		
SNP	CHR	BP	EA	P	BETA	SE	P	BETA	SE
rs12131421	1	5472626	A	2.45E-07	-0.327	0.066	9.3E-01	0.006	0.063
rs977395	3	3968679	T	9.40E-07	-0.242	0.048	1.2E-01	-0.091	0.058
rs75824901	17	6804102	A	9.56E-07	-0.624	0.140	6.9E-01	0.047	0.118
rs10988763	9	102076659	A	1.03E-06	-0.131	0.027	5.2E-01	0.019	0.029
rs77382989	13	39848089	G	1.14E-06	-0.788	0.183	7.9E-01	0.038	0.142
rs72945436	6	106034914	C	1.32E-06	0.217	0.044	9.3E-01	-0.005	0.051
rs1419863	7	34537783	A	1.52E-06	0.123	0.025	4.9E-01	0.019	0.028
rs12672727	7	139354457	C	1.68E-06	-0.254	0.055	7.2E-01	-0.020	0.054
rs72923554	2	193518421	T	1.72E-06	0.281	0.056	7.9E-02	-0.123	0.070
rs7546521	1	229103599	G	1.88E-06	0.240	0.049	9.7E-01	0.002	0.059
rs74702184	3	65627458	T	1.97E-06	0.175	0.036	1.2E-01	0.066	0.042
rs7209250	17	71029699	C	9.07E-02	0.058	0.035	8.6E-08	0.194	0.035
rs4782356	16	88597767	G	1.48E-01	-0.046	0.032	1.5E-06	0.165	0.034
rs34455020	6	168837096	A	4.38E-01	0.098	0.126	1.3E-06	0.558	0.106
rs7841134	8	13603585	T	7.93E-01	0.015	0.057	1.0E-06	-0.334	0.072

Next, separate GWAS were conducted for the different sources of clinical diagnoses of fibromyalgia, namely hospital (ICD based) and primary care (read codes based) to further delve into case ascertainment-based heterogeneity within the clinically diagnosed case group (Figure 8). Q-Q plots and genomic inflation factors were as expected: λ_{GC} = 1.018 (ICD) and λ_{GC} = 1.032 (primary care) (Figure 9). Several lead

SNPs could be observed in the primary care GWAS, namely on chromosomes 3, 5 and 9. No lead SNPs were identified in the hospital diagnosed fibromyalgia only GWAS, potentially due to low case numbers (Figure 8, Table 9).

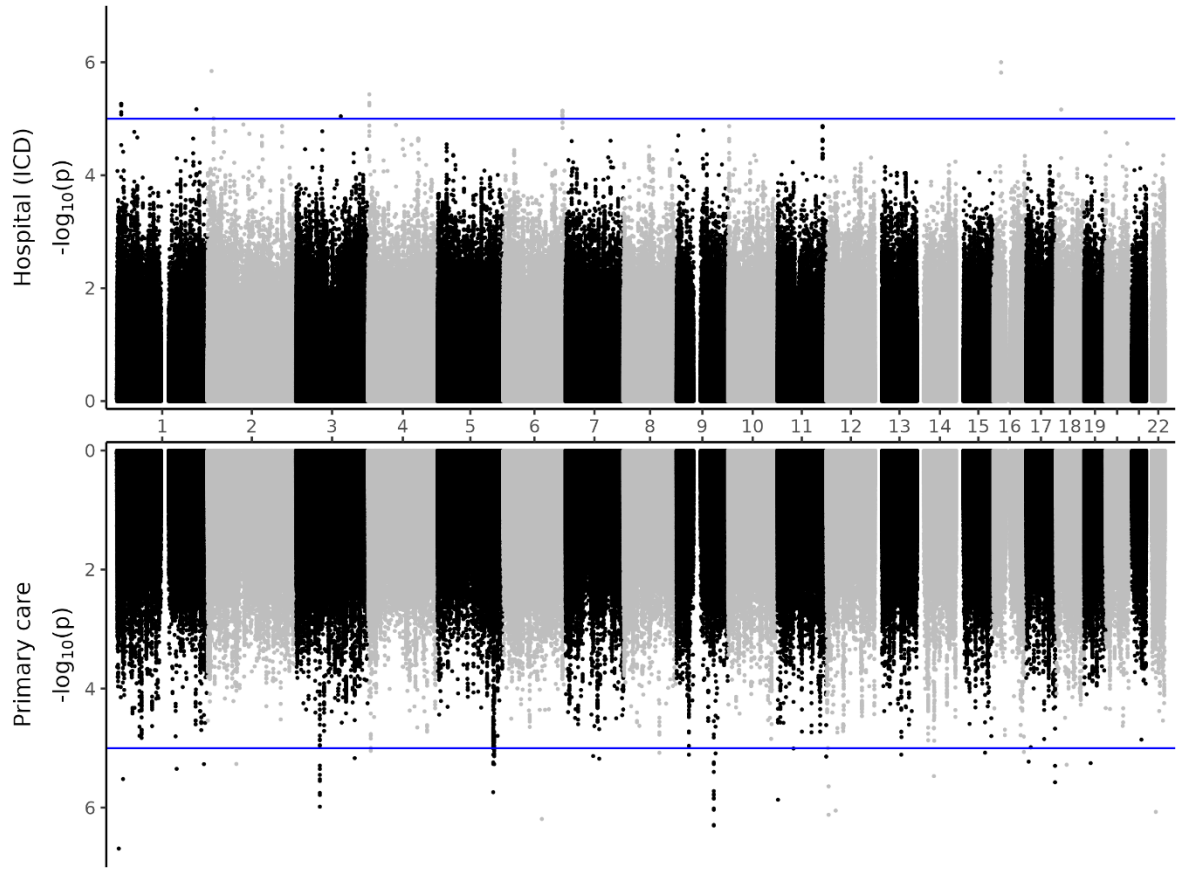


Figure 8. Miami plot of ICD-based fibromyalgia GWAS (top) versus primary care fibromyalgia GWAS (bottom). The blue line indicates suggestive level of significance ($p<10^{-5}$).

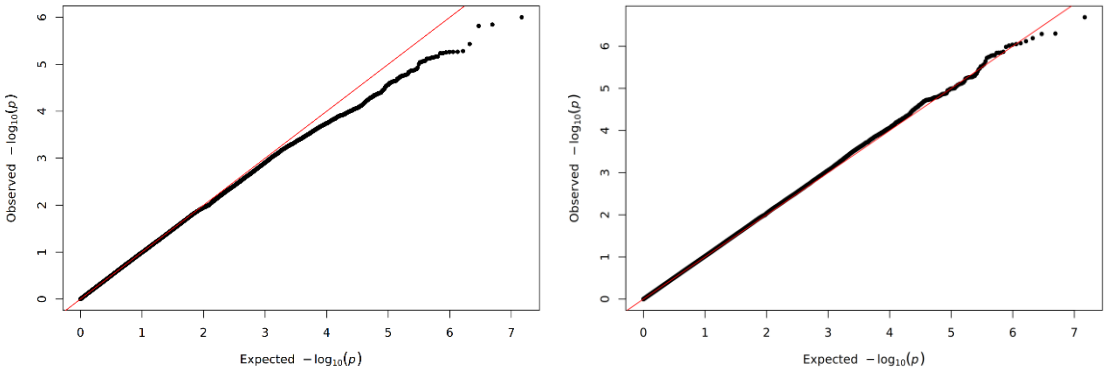


Figure 9. Q-Q plots of ICD-based fibromyalgia GWAS (left) and primary care fibromyalgia GWAS (right).

The lack of nominal or significant associations in the ICD-based diagnosis of fibromyalgia GWAS is not surprising given that there are just over 500 ICD diagnosed fibromyalgia cases in the UK Biobank. However, since none of the top signals in the primary care data were at least nominally significant in the ICD GWAS, there is the possibility the ICD cohort and the primary care cohort presenting different ends of the spectrum for fibromyalgia severity or there are issues around clinical diagnosis of the condition between the two categories (Table 9). The FinnGen cohort defined fibromyalgia cases utilising ICD codes only. I have compared the effect sizes and p-values of lead SNPs with ICD-based fibromyalgia GWAS from FinnGen (Table 9). Surprisingly, neither UK Biobank fibromyalgia ICD GWAS nor FinnGen fibromyalgia GWAS lead signals are shared with each other. However, as FinnGen utilises ICD codes to identify the fibromyalgia cohort, the rest of the analysis will focus on the ICD cohort of UK Biobank.

Table 9. Lead SNPs ($p<10^{-6}$) of fibromyalgia primary care GWAS, ICD GWAS and FinnGen GWAS.

					PRIMARY CARE			ICD			FINNGEN		
SNP	CHR	BP	EA	EAF	P	BETA	SE	P	BETA	SE	P	BETA	SE
rs12131421	1	5472626	A	0.051	2.06E-07	-0.361	0.073	6.84E-02	-0.316	0.173	1.93E-02	0.149	0.064
rs446207	9	102077408	A	0.606	5.01E-07	0.146	0.029	8.69E-01	0.013	0.079	2.79E-01	0.033	0.031
rs72945436	6	106034914	C	0.103	6.45E-07	0.241	0.047	8.17E-01	-0.033	0.141			
rs116874789	12	6034400	C	0.060	7.60E-07	-0.318	0.067	4.79E-01	0.114	0.160	6.76E-01	0.033	0.079
rs6005014	22	26718993	G	0.041	8.49E-07	0.319	0.062	8.96E-01	0.025	0.192	7.59E-01	0.026	0.083
rs7961473	12	25513333	T	0.831	8.92E-07	-0.188	0.038	7.13E-01	-0.040	0.109	1.23E-01	0.070	0.046
rs11714831	3	65659142	G	0.174	1.04E-06	0.174	0.035	4.38E-01	-0.079	0.102	3.68E-01	0.034	0.038
rs141412360	11	1847486	A	0.019	1.36E-06	0.444	0.086	7.25E-01	0.100	0.285	9.18E-02	-0.168	0.100
rs114162268	5	153120461	T	0.015	1.82E-06	0.480	0.094	6.96E-01	-0.125	0.320			
rs78847386	16	21165368	A	0.030	9.49E-01	0.005	0.084	9.98E-07	0.868	0.156	2.22E-01	0.099	0.081
rs114790411	2	11247351	A	0.031	9.12E-01	-0.009	0.082	1.43E-06	0.831	0.152	2.55E-02	0.242	0.109
rs6045664	20	1940145	C	0.947	6.80E-01	0.018	0.045	6.41E-01	0.058	0.124	1.65E-07	-0.320	0.061
rs12481416	20	5587336	T	0.069	9.19E-01	-0.006	0.056	1.67E-01	-0.217	0.157	2.80E-07	-0.329	0.064
rs147764067	16	18057228	G	0.030	2.73E-01	0.108	0.099	8.48E-01	-0.052	0.270	1.42E-06	-0.485	0.101
rs7236724	18	55652171	C	0.618	1.95E-01	0.038	0.029	9.53E-01	-0.005	0.080	1.95E-06	-0.142	0.030

3.3.5. Meta-analysis of fibromyalgia

Meta-analysis between UK Biobank (N cases = 532, N controls = 240,656) and FinnGen (N cases = 3166, N controls = 470,515) was conducted utilising the ICD-defined case selection. Lambda GC for the meta-analysis was 1.013. The meta-analysis revealed one lead signal, rs34323745, was identified on chromosome 7 in the intronic region of *GPNMB*, that almost reached genome-wide significance (OR = 1.19, 95% CI: [1.13, 1.26], $p = 7.27 \times 10^{-8}$) in the meta-analysis (Figure 10). There are also 6 nominal signals passing the $p < 5 \times 10^{-6}$: (1) lead rs114790411 on chromosome 2 ($p = 7.42 \times 10^{-7}$), (2) lead rs3801680 on chromosome 7 ($p = 4.88 \times 10^{-7}$), (3) lead rs11721715 on chromosome 4 ($p = 4.29 \times 10^{-6}$), (4) lead rs12524339 on chromosome 6 ($p = 2.78 \times 10^{-6}$). (5) lead rs10796307 on chromosome 10 ($p = 1.42 \times 10^{-6}$), (6) on chromosome 20 rs12481416 ($p = 2.54 \times 10^{-7}$).

The lead nominal SNPs from meta-analysis were annotated utilising large scale eQTL maps from GTEx, eQTLGen and BRAINEAC. The top SNP on chromosome 7,

rs34323745, is an eQTL in several tissues potentially relevant to fibromyalgia: blood (eQTLGen)[101], frontal cortex, putamen, temporal cortex and average expression across ten brain regions (BRAINEAC)(Table 10)[102].

The lead signal was mapped to several genes, according to eQTL mapping: *KLHL7*, *GPNMB*, *FAM126A*, *IGF2BP3*, *MALSU1*, *TOMM7* and *NUPL2*. In particular, expression of *GPNMB* is regulated by rs34323745 in brain tissues. GTEx revealed that rs34323745 is also an eQTL in pituitary gland, cerebellum, and skeletal muscle (Table 11).

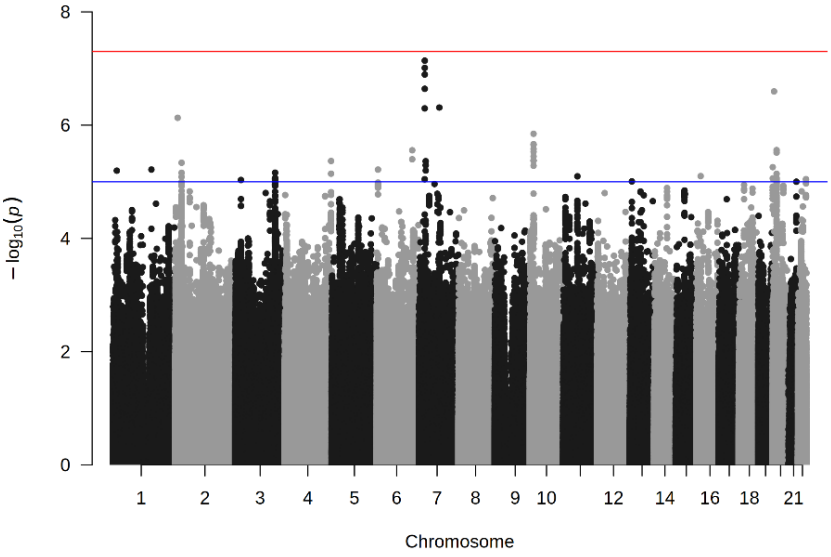


Figure 10. Manhattan plot for meta-analysis of fibromyalgia. Blue line represents a nominal threshold of significance ($p<10^{-5}$). Red line represents genome-wide significance ($p<5x10^{-8}$).

To test the hypothesis that the differences between primary care GWAS and ICD GWAS in UK Biobank were due to the low number of cases in the latter cohort, I compared the GWAS results from primary care to the meta-analysis, which added an additional ~3,000 cases of fibromyalgia. Even with this boost in power, there are no shared lead signals between the two GWAS (Figure 11).

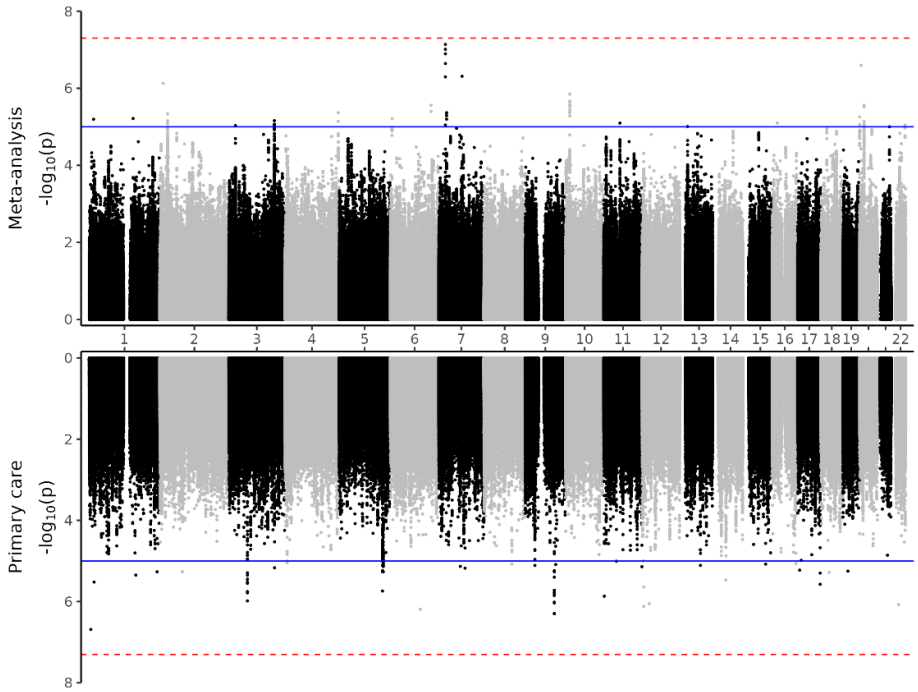
Table 10. Expression quantitative trait loci for the rs34323745, the top GWAS signal in the fibromyalgia GWAS meta-analysis. Results were obtained from GTEx Portal (v8)[75].

Gencode Id	Gene Symbol	SNP Id	P-Value	NES	Tissue
ENSG00000230658.1	KLHL7-AS1	rs34323745	1.20E-08	-0.43	Nerve - Tibial
ENSG00000136243.16	NUPL2	rs34323745	1.10E-07	0.52	Brain - Cortex
ENSG00000230658.1	KLHL7-AS1	rs34323745	4.10E-07	-0.59	Brain - Cortex
ENSG00000136235.15	GPNMB	rs34323745	4.20E-07	0.57	Brain - Nucleus accumbens (basal ganglia)
ENSG00000230658.1	KLHL7-AS1	rs34323745	4.10E-06	-0.27	Muscle - Skeletal
ENSG00000230658.1	KLHL7-AS1	rs34323745	1.10E-05	-0.46	Pituitary
ENSG00000230658.1	KLHL7-AS1	rs34323745	1.60E-05	-0.54	Brain - Nucleus accumbens (basal ganglia)
ENSG00000136235.15	GPNMB	rs34323745	1.90E-05	0.39	Brain - Cortex
ENSG00000196683.10	TOMM7	rs34323745	1.90E-05	0.13	Nerve - Tibial
ENSG00000136243.16	NUPL2	rs34323745	4.10E-05	0.33	Brain - Cerebellum

Table 11. Expression quantitative trait loci for rs34323745, the lead GWAS signal in the fibromyalgia GWAS meta-analysis. Results were obtained from FUMA, using eQTLGen and BRAINEAC databases. False Discovery Rate (FDR) was calculated by FUMA.

Gene	P	FDR	Tissue	Database
KLHL7	5.35E-25	0	eQTLGen_cis_eQTLs	eQTLGen
GPNMB	4.57E-139	0	eQTLGen_cis_eQTLs	eQTLGen
NUPL2	2.27E-148	0	eQTLGen_cis_eQTLs	eQTLGen
GPNMB	1.81E-03	0.029	Frontal Cortex	BRAINEAC
GPNMB	4.24E-05	0.003	Putamen	BRAINEAC
GPNMB	3.47E-04	0.006	Temporal cortex	BRAINEAC
GPNMB	1.81E-03	0.029	average across all 10 brain tissues	BRAINEAC

Figure 11. Miami plot of GWAS meta-analysis of fibromyalgia (top) versus primary care GWAS (bottom). Blue line represents a nominal threshold of significance ($p<10^{-5}$). Red, dotted line represents genome-wide significance ($p<5x10^{-8}$).



3.4. Discussion

In this chapter, I have completed a sex stratified GWAS of fibromyalgia in UK Biobank, which revealed different male and female signals; compared different methods for conducting GWAS and selected REGENIE as the most appropriate for further analysis; conducted sensitivity analysis and showed different fibromyalgia case definitions have dramatically different genetic signals, and lastly, I have conducted a meta-analysis of fibromyalgia GWAS using FinnGen. The meta-analysis has revealed that rs34323745 in the *GPNMB* locus was almost genome-wide significant ($p = 7.269 \times 10^{-8}$).

3.4.1. Sex-specific genetic signals for Fibromyalgia in UK Biobank

Sex-stratified GWAS of fibromyalgia in UK Biobank has highlighted different male- and female-specific signals. Although not statistically significant, the most prominent signal in female and sex-combined GWAS is an intronic SNP in *COMMD1* gene.

This gene encodes for a protein involved in copper metabolism, and in protein degradation of *NF-κB*, which results in downregulation of TNF-induced inflammation[103,104]. Moreover, *COMMD1* has been shown to have an anti-inflammatory role in murine myeloid cells and its expression is decreased in IBD patients[105].

The top male signal is intronic to *ARSG* gene, a lysosomal sulfatase with no obvious connection to pain pathways. The Manhattan plot indicates the signal might be a rare variant or a spurious association, because the SNPs in proximity do not have a similarly low p-value. Indeed, variant rs75888727 has a MAF of 1%. Given the low number of cases in males, the UK Biobank and FinnGen male to female ratio is heavily skewed towards a female preponderance of fibromyalgia. However, a recent study found evidence of fibromyalgia being underdiagnosed in males[47].

3.4.2. Fibromyalgia case definition: Heterogeneity and possible data discrepancies

Sensitivity analyses have revealed there is heterogeneity between the different fibromyalgia case definitions. This suggests, firstly, that there are important differences between the self-reported data and the clinically diagnosed cohort. Cohort characteristics appear similar between the two groups, except for overall health rating at the time of recruitment and educational attainment, which varied between the two groups: ICD versus Primary care and Clinical diagnosis versus Self-reported fibromyalgia. ICD-based diagnosis of fibromyalgia and overall clinical diagnosis of fibromyalgia cohorts appear to have worse self-reported overall health ratings. This could indicate the two cohorts may have more comorbidities. However, this difference may also not be clinically relevant as we do not have this information at the time of diagnosis, but rather at the time of recruitment. A few potential

explanations for this discrepancy between cohorts are the self-reports are subject to recall bias, under- or overrepresentation of certain subpopulations in the case pool, etc. There is also the possibility some of the self-reports are not true fibromyalgia cases, however both questions in the baseline and follow up questionnaire ask specifically if the person was previously diagnosed by a doctor. Another surprising difference concerns the genetic heterogeneity between hospital diagnosed patients, referred to as the ICD cohort, and primary care cases. Most clinical records of fibromyalgia are from primary care. Although general practitioners are not likely to diagnose fibromyalgia themselves, the diagnostic code is potentially a consequence of a letter from a specialist. It is possible a hospital record of fibromyalgia, which has been performed by a rheumatologist, might indicate a more severe case. In this case, it is possible the patients with hospital diagnosis of fibromyalgia have other chronic pain conditions. However, neither hospital nor primary care records are complete, and it is likely diagnoses were lost for many of the patients. For primary care, records were only retrieved for two thirds of participants, however the coverage and completeness of records is not known. When participants move, their records may not be updated.

Another possibility is that the diagnostic criteria between ICD, primary care and even FinnGen GWAS changed between the different categories of patients. Both ICD10 data and American College of Rheumatology criteria date back to the early 1990s, with updated criteria being published by American College of Rheumatology in 2010/2011 and 2016. The biggest difference in diagnostic criteria is that the latest do not use a tender point site count, but rather use a widespread pain index and a symptom severity scale[49]. It could be hypothesised that either the different case definitions represent different severity degrees on a continuous “fibromyalgia-ness”

scale, or rather different musculoskeletal disorders are being misdiagnosed as fibromyalgia. The concept of “fibromyalgia-ness” was introduced by Wolfe in 2019 and it involves rethinking of this syndrome as a continuous distribution of symptoms, such as fatigue and non-articular pain sites combined into a symptom intensity score[106].

3.4.3. Functional annotation of nominal associations with fibromyalgia

There was one lead SNP which almost reached genome-wide significance in the ICD-coded fibromyalgia GWAS meta-analysis. The SNP, rs34323745, has been mapped to multiple genes using eQTL analysis. It is possible it regulates the expression of many genes in a tissue-dependent manner. Of interest for fibromyalgia is *GPNMB*, a gene expressed in many brain tissues and microglia, involved in modulating neuroinflammation. It is thought to mainly have an anti-inflammatory effect. Moreover, its expression is increased in many neurodegenerative conditions[107]. In peripheral tissues, its expression is increased in macrophages involved in pro-inflammatory conditions[108].

Given the first GWAS meta-analysis results for fibromyalgia in this chapter showed a promising lead SNP and the epidemiological link of increased co-morbidity with endometriosis, in the next chapter, I will investigate the genetic correlation between the meta-analysis of fibromyalgia and the largest published meta-analysis of endometriosis, and then perform multi trait association analysis of fibromyalgia and endometriosis in order to boost the power of detection of genome-wide significant signals in fibromyalgia.

Chapter IV. Genetic sharing of endometriosis and fibromyalgia

4.1. Introduction

Increased comorbidity between endometriosis and other chronic pain conditions is well recognised, with mechanisms widely accepted to include central and peripheral sensitisation[59,61,62,66]. People with endometriosis can experience pain in different ways, most commonly generalised pain in the pelvis and bowels[59,109]. Although endometriosis lesions were thought to be the cause of pain, their removal does not improve pain in about a quarter of patients[110]. Endometrial fragments are thought to flow back into the peritoneum due to retrograde menstruation, where they attach, create blood vessels, and become innervated, which ensures their survival[58,59,111]. The innate immune system is recruited to the site and creates a pro-inflammatory environment to clear out the fragment[112]. Failing to clear them, inflammation does not subside, and sensory nerves undergo changes due to persistent activation. Long-lasting nociceptive signalling from the lesions strengthens the communication between periphery and central nervous system and lowers the threshold for activation[59,80,113,114]. Lastly, some women experience chronic generalized hypersensitivity across visceral organs, due to potential cross talk between reproductive tract nerves (splanchnic, pelvic, and pudendal) with nearby organs[115]. Similarly, fibromyalgia-associated pain mechanisms are thought to be peripheral and central sensitisation[116]. The increased comorbidity between the two conditions has been demonstrated previously[65,66]. People with endometriosis are at higher risk of subsequently developing fibromyalgia, according to a Swedish study by Pardo *et al.* (2019) (incident rate ratios = 2.83, 95% CI: 1.96 – 4.08)[117]. Another study showed people with endometriosis to have higher rates of fibromyalgia than the general female US prevalence (5.9 versus 3.4%, $P < 0.0001$)[118].

The largest GWAS meta-analysis of endometriosis, conducted by Rahmioglu *et al.*, (2023), contains 24 studies, including UK Biobank, totalling 60,674 cases and 701,926 controls) of European (98%—Europe, USA, Australia) and East Asian (2%—Japan) ancestry[40]. They have shown that endometriosis is genetically correlated with migraine (rg = 0.29), dorsalgia (rg = 0.45), multisite chronic pain (rg = 0.43) and chronic back pain (rg = 0.33) in the UK Biobank. Therefore, endometriosis is genetically correlated with musculoskeletal chronic pain traits in the UK Biobank.

These previous studies point to the possibility of genetic sharing between endometriosis and fibromyalgia. However, the genetic sharing between fibromyalgia and endometriosis has not been investigated in depth. In this chapter, I explore the genetic relationship between endometriosis and fibromyalgia utilising the GWAS meta-analysis of fibromyalgia results from the previous chapter with the GWAS meta-analysis of endometriosis from Rahmioglu *et al.*, (2023) to discover further novel genetic variants for fibromyalgia[40].

4.2. Methods

4.2.1. Meta-analysis of endometriosis GWAS

The endometriosis meta-analysis conducted by Rahmioglu *et al.*, (2023) included 24 studies, 12 of which had surgically confirmed endometriosis cases and 15 which had more than 300 cases[40]. Summary statistics are publicly available for a total of 21,779 European ancestry cases and 449,087 European ancestry controls. Meta-analysis was conducted for 7,542,693 SNPs under a fixed-effects model with inverse variance weighting using METAL[98]. Here I utilised the European ancestry based GWAS meta-analysis summary statistics.

4.2.2. Meta-analysis of fibromyalgia GWAS

I utilised the European ancestry fibromyalgia GWAS meta-analysis I conducted in Chapter III including 3,698 cases with ICD-based diagnosis and 699,009 controls from UK Biobank and FinnGen datasets. FinnGen biobank uses ICD-10 diagnosis to define fibromyalgia cases. UK Biobank reports varied sources for fibromyalgia case definition. However, due to heterogeneity of these different sources, only ICD-10 cases were selected for the meta-analysis, in line with the FinnGen case definition (see Chapter III).

4.2.2. SNP-heritability estimation and genetic correlation using LD-score regression

Genetic correlation between fibromyalgia meta-analysis and endometriosis meta-analysis was conducted using LD-score regression (see Chapter III, Methods, SNP-heritability estimation using LD-score regression)[99]. Both summary statistics were prepared for analysis using the munge_sumstats.py function of LDSC. A statistically significant threshold of 0.05 was considered.

SNP-heritability scores for fibromyalgia were converted from the observed scale to the liability scale by providing the --sample-prev (disease prevalence in the sample) and --pop-prev (disease prevalence in the population) flags in the ldsc.py function. Prevalence for fibromyalgia in the general population was estimated at 2%, from the scientific literature, and in the meta-analysis at 0.5%[47,50]. Prevalence of endometriosis in the general population was estimated at 8%, and in the meta-analysis at 8%[40,58].

4.2.3. Multi-trait association analysis

A joint genetic analysis of endometriosis and fibromyalgia was conducted using multi-trait analysis of GWAS (MTAG)[119]. MTAG assumes genetic correlation between the traits investigated and allows better effect size estimates by improving power of detection. SNPs were first filtered out from both datasets if they had MAF = <0.01, were mapped to the same chromosomal positions, had different alleles between datasets, had MAF differences of >0.2 between datasets or non-autosomal location. Only SNPs that were shared between the GWAS analyses were utilized in the analysis. Z scores (log(OR/SE)) were computed for all SNPs. After variant filtering, a total of 1,104,043 variants were included in the analysis. Sample overlap between GWAS results was accounted for using bivariate LD score regression. In the results, MTAG provided trait-specific effect estimates for each SNP, and the resulting p-values that can be interpreted and used like those in single-trait GWAS. Lastly, MTAG calculated maxFDR for each trait.

4.2.4. Biology of fibromyalgia MTAG variants

FUMA takes GWAS summary statistics as input and outputs functional annotations of variants[73]. It was also utilised to identify the lead SNPs at the selected $p < 10^{-5}$ threshold. In terms of functional annotation, the lead variants were mapped to genes positionally, then, in a separate analysis, they were mapped according to whether they were eQTLs for particular genes in the largest blood eQTL database, eQTLGen, and the largest brain eQTL database, BRAINEAC[101,102]. Tissue gene expression information was obtained from GTEx version 8[75]. FUMA was utilised to compute tissue-based differentially expressed gene sets and to perform gene enrichment analysis. Significant gene sets were identified if the Bonferroni corrected p-value was less than 0.05.

4.3. Results

4.3.1. Fibromyalgia and endometriosis are strongly genetically correlated

LD-score regression analysis utilising endometriosis and fibromyalgia GWAS meta-analysis results estimated the conditions are strongly genetically correlated (r_g = 0.46, Z-score: 8.05, p = 8.52 x 10^{-16}). The SNP-heritability for fibromyalgia was estimated as 14.52% and for endometriosis as 26.98%.

4.3.2. Multi trait association analysis for fibromyalgia

Genomic inflation factor for the MTAG of fibromyalgia was = 1.059 (Figure 12). Two genome-wide association signals were detected for fibromyalgia following MTAG analysis of endometriosis and fibromyalgia meta-analyses: rs17082358, an intronic SNP within *SYNE1* gene on chromosome 6 (OR: 0.98, 95% CI: [0.985, 0.992], p = 1.49 x 10^{-11}) and rs13432756, an intronic SNP within *GREB1* gene on chromosome 2 (OR = 1.013, 95% CI: [1.009, 1.017], p = 6.39 x 10^{-9}) (Figures 13, 14). They are nominally associated with fibromyalgia in the univariate GWAS analyses; however, the strength of the association appears to be driven by the very low p-value in the univariate GWAS of endometriosis (Table 12). In addition, there are 22 lead SNPs (p < 10^{-5}) for fibromyalgia that are mostly nominally significant (p < 10^{-3}) in the single trait GWAS of endometriosis and fibromyalgia (FinnGen) (Table 12). Most lead SNPs do not appear to be shared by the univariate GWAS of fibromyalgia in UK Biobank. Therefore, most lead SNPs in the MTAG of fibromyalgia are driven by the GWAS of fibromyalgia in FinnGen. The mean χ^2-statistic for endometriosis was 1.139 and for fibromyalgia was 1.054. Max FDR for endometriosis was 0.033 and for fibromyalgia was 0.145. Potential false positives are rs72478520, rs74485684 and rs28613416, because their p-values are not at least nominally significant in either univariate GWAS of fibromyalgia (Table 12). Several SNPs might be non-null for fibromyalgia

and null for endometriosis based on high p-values for endometriosis MTAG and univariate GWAS meta-analysis (rs11678188 on chr 2, rs3801680 on chr7, rs1021715 on chr 5, rs2984646 on chr 5, rs34185333 on chr 7, rs12524339 on chr 6, rs17050273 on chr 2).

Variants of interest were not nominally significant in fibromyalgia MTAG (rs11172113 on chr 12, rs6478241 on chr 9, rs7523086 on chr 1). I then investigated the loci of the genes of interest: *NGF*, *ASTN2* and *LRP1* and identified the variants with the lowest p-values: rs150043534 on chr 1 (p-value = 0.025), rs923601 on chr 9 (p-value = 0.002), and rs1131514 on chr 12 (p-value = 0.0023). These p-values are not significant and do not provide evidence for a link between these genes and fibromyalgia. It is possible that there are nominally significant variants further away from the gene locus that regulate gene expression of the genes of interest.

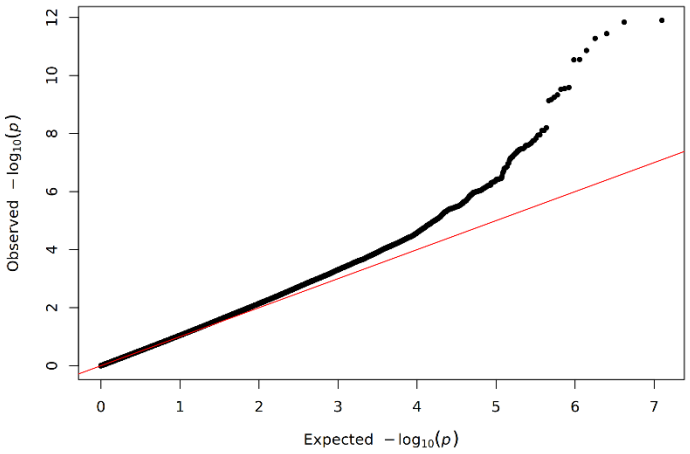


Figure 12. Q-Q plot of multi trait association analysis of fibromyalgia.

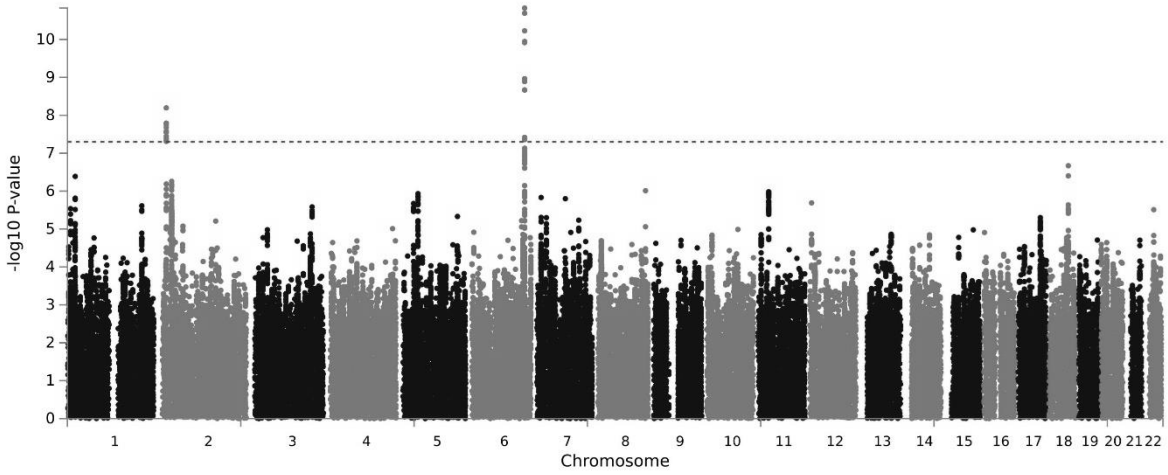


Figure 13. Manhattan plot of multi trait association analysis of fibromyalgia. Dotted line represents genome wide significance ($p<5x10^{-8}$).

Table 12. Lead SNPs (p < 10-5) from multi-trait association analysis of fibromyalgia (MTAG) versus univariate GWAS of endometriosis (ENDO) and fibromyalgia (meta-analysis, UK Biobank and FinnGen). EA, effect allele; EAF, effect allele frequency.

SNP	CHR	BP	EA	EAF	FIBRO MTAG				ENDO MTAG				ENDO META-ANALYSIS				FIBRO META-ANALYSIS				FIBRO UK BIOBANK				FIBRO FINNGEN			
					P	BETA	SE		P	BETA	SE		P	BETA	SE		P	BETA	SE		P	BETA	SE		P	BETA	SE	
rs17082358	6	152545865	T	0.77	1.49E-11	-0.012	0.002		5.57E-22	-0.060	0.006		4.01E-19	-0.127	0.014		6.19E-03	0.092	0.034		5.53E-01	0.058	0.099		5.75E-03	0.096	0.035	
rs13432756	2	11660615	T	0.12	6.39E-09	0.013	0.002		4.22E-15	0.055	0.007		6.42E-13	0.117	0.016		8.57E-03	-0.113	0.043		4.63E-03	0.294	0.100		1.22E-01	0.072	0.047	
rs7236724	18	55652171	A	0.38	2.15E-07	0.008	0.001		1.48E-04	0.020	0.005		5.43E-03	0.034	0.012		1.32E-05	-0.125	0.029		9.53E-01	-0.005	0.080		1.95E-06	-0.142	0.030	
rs72478520	1	22489567	T	0.20	4.10E-07	0.009	0.002		2.40E-19	0.056	0.006		3.86E-18	0.126	0.015		3.34E-01	0.138	0.035		2.11E-01	-0.126	0.101		1.26E-01	0.056	0.037	
rs11678188	2	2724832	T	0.34	5.53E-07	-0.008	0.002		2.43E-03	-0.015	0.005		4.46E-02	-0.024	0.012		4.62E-06	0.138	0.030		8.06E-02	0.139	0.080		1.41E-05	0.138	0.032	
rs11166700	8	138295559	T	0.39	9.76E-07	0.007	0.001		9.61E-04	0.017	0.005		1.92E-02	0.027	0.012		1.94E-05	-0.123	0.029		3.03E-01	-0.082	0.080		1.83E-05	-0.129	0.030	
rs74485684	11	30242287	T	0.82	1.04E-06	0.009	0.002		8.04E-23	0.066	0.007		3.77E-22	0.153	0.016		8.35E-01	-0.008	0.037		6.53E-02	-0.197	0.107		6.43E-01	0.018	0.038	
rs145600094	5	43689656	T	0.05	1.17E-06	0.015	0.003		1.42E-04	0.036	0.009		3.55E-03	0.066	0.023		9.04E-05	-0.231	0.059		2.91E-03	0.389	0.123		4.71E-03	0.184	0.065	
rs79573358	7	12052387	A	0.05	1.47E-06	0.016	0.003		2.30E-05	0.054	0.013		6.65E-04	0.099	0.029		3.23E-04	-0.222	0.062		4.99E-01	0.139	0.205		2.62E-04	0.230	0.063	
rs3801680	7	82014682	T	0.86	1.59E-06	0.010	0.002		6.18E-02	0.014	0.007		4.56E-01	0.013	0.018		4.89E-07	-0.213	0.042		7.29E-03	-0.325	0.125		6.21E-06	-0.199	0.044	
rs4930767	12	5597381	T	0.48	2.05E-06	0.007	0.001		8.51E-04	0.016	0.005		1.55E-02	0.028	0.011		4.90E-05	-0.115	0.028		3.50E-02	-0.163	0.077		2.92E-04	-0.107	0.030	
rs1021715	5	30906612	T	0.19	2.12E-06	-0.009	0.002		2.68E-03	-0.019	0.006		4.05E-02	-0.030	0.014		2.07E-05	0.157	0.037		5.98E-03	-0.286	0.106		2.76E-04	-0.140	0.038	
rs10735513	1	214520268	T	0.23	2.44E-06	-0.008	0.002		1.02E-07	-0.032	0.006		3.13E-06	-0.066	0.014		5.38E-03	0.095	0.034		3.93E-01	-0.083	0.097		7.54E-04	0.120	0.036	
rs6773027	3	165567905	A	0.25	2.61E-06	0.008	0.002		6.99E-04	0.019	0.006		1.26E-02	0.032	0.013		7.42E-05	-0.127	0.032		6.51E-02	-0.160	0.087		2.92E-04	-0.122	0.034	
rs2765513	1	8987597	T	0.22	2.91E-06	-0.008	0.002		7.43E-04	-0.020	0.006		1.31E-02	-0.035	0.014		7.97E-05	0.137	0.035		1.35E-02	-0.234	0.095		8.03E-04	-0.122	0.036	
rs131184	22	29681216	A	0.89	3.10E-06	-0.011	0.002		9.59E-04	-0.024	0.007		1.61E-02	-0.040	0.017		7.12E-05	0.171	0.043		5.40E-01	0.068	0.112		3.50E-05	0.188	0.046	
rs2984646	5	157594700	T	0.74	4.67E-06	-0.007	0.002		1.49E-03	-0.017	0.005		2.20E-02	-0.030	0.013		8.15E-05	0.125	0.032		1.78E-01	0.117	0.087		1.46E-04	0.126	0.033	
rs28613416	17	64058429	T	0.10	4.99E-06	-0.011	0.002		2.55E-08	-0.054	0.010		6.43E-07	-0.113	0.023		1.45E-02	0.118	0.048		7.20E-01	-0.055	0.152		1.18E-02	-0.125	0.050	
rs34185333	7	27239193	T	0.05	5.02E-06	0.015	0.003		3.19E-03	0.043	0.014		4.24E-02	0.069	0.034		4.83E-05	-0.256	0.063		4.28E-02	0.487	0.241		1.70E-04	0.239	0.064	
rs41705	7	120645068	A	0.23	5.86E-06	0.008	0.002		5.12E-05	0.024	0.006		1.03E-03	0.044	0.013		8.46E-04	-0.111	0.033		2.01E-01	-0.120	0.094		1.58E-03	-0.110	0.035	
rs12524339	6	143739719	T	0.97	6.02E-06	0.019	0.004		6.31E-02	0.022	0.012		4.11E-01	0.024	0.029		2.78E-06	-0.461	0.098		1.01E-03	-0.753	0.253		8.24E-05	-0.409	0.104	
rs72862536	2	154307813	T	0.04	6.19E-06	-0.018	0.004		2.19E-05	-0.045	0.011		4.69E-04	-0.090	0.026		1.33E-03	0.256	0.080		2.91E-02	-0.368	0.169		1.10E-02	-0.224	0.088	
rs17050273	2	59810624	A	0.19	8.41E-06	-0.008	0.002		2.23E-02	-0.014	0.006		1.86E-01	-0.019	0.015		1.48E-05	0.159	0.037		5.69E-01	-0.057	0.100		5.82E-06	-0.175	0.039	
rs79851790	4	180199103	A	0.06	9.77E-06	-0.014	0.003		7.79E-04	-0.032	0.009		1.11E-02	-0.061	0.024		2.90E-04	0.248	0.068		8.82E-01	-0.022	0.149		4.21E-05	-0.307	0.075	

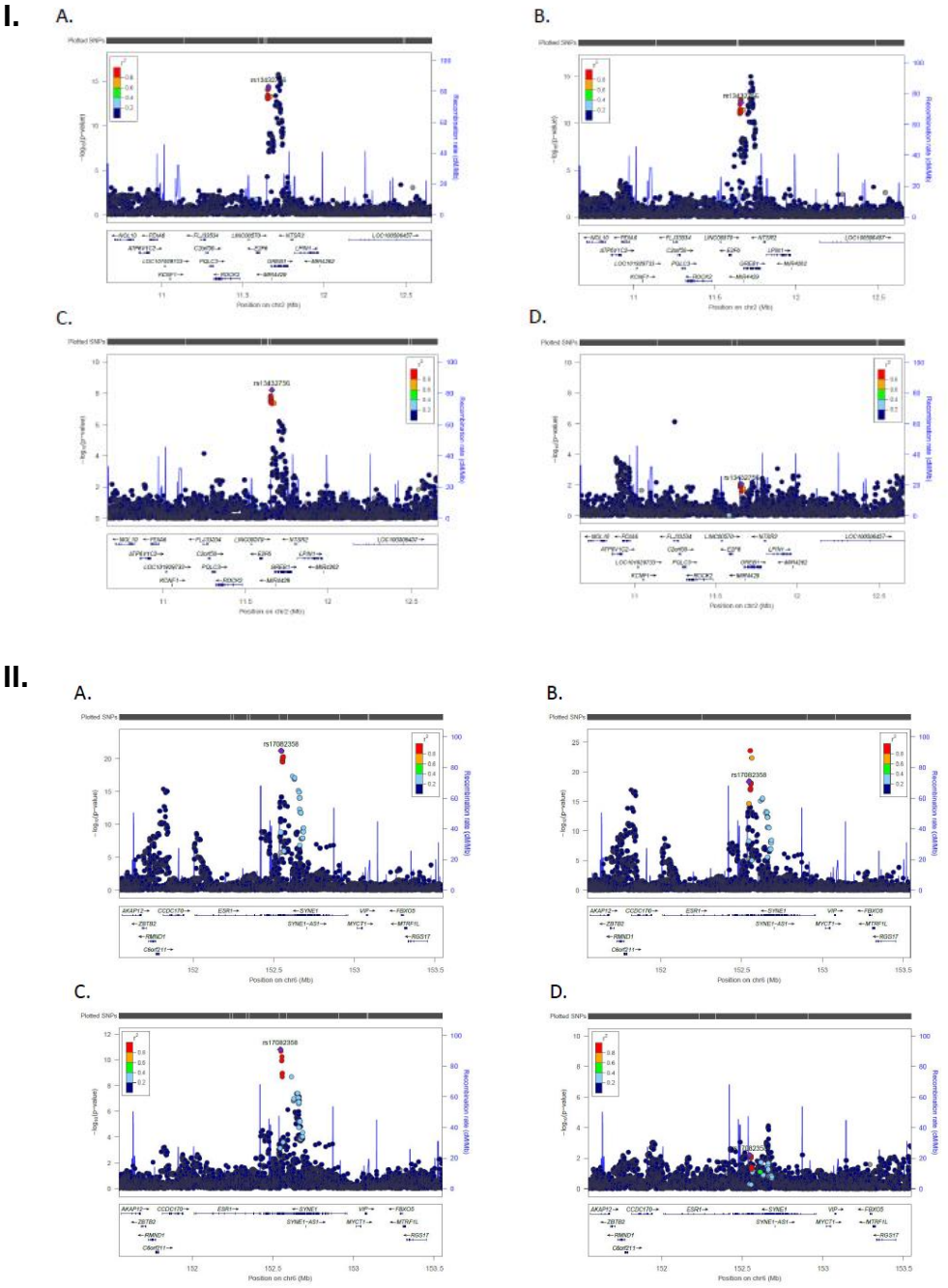


Figure 14. Regional association plots of SYNE1/6q25.1 (I) and GREB1/2p25.1 (II) loci, the top signals identified from the multi trait association analysis (MTAG) of fibromyalgia and endometriosis (purple dot). A: Endometriosis MTAG; B: Endometriosis meta-analysis; C: Fibromyalgia MTAG; D: Fibromyalgia meta-analysis.

4.3.3. Biology of lead signals for fibromyalgia MTAG

Lead SNPs are located mostly intergenic or intronic. They form distinct clusters depending on the heatmap created by FUMA using tissue expression from GTEx (v8)[73,75]. When genes were mapped positionally, several clusters could be observed (Figure 15). One cluster represents genes highly expressed everywhere. They regulate processes like anchoring the cytoskeleton (*MAPRE3*, *GAS2L1*), apoptosis (*TMEM214*), hormone signalling (*PREB*), post-translational modifications (*OST4*), transcription regulation (*EWSR1*), translation of mRNAs (*PAIP1*) and vesicle transport (*AP1B1*)[85,120–126]. Other highly expressed genes in most of the 54 tissue types are *RHBDD3*, involved in liver development and proteolysis, *NNT*, important for NADPH production, *SYNE1*, involved in actin skeleton organisation, *SMYD2*, a lysine methylase, *AGBL5*, a metallocarboxypeptidase and *ARL14EP*, which controls the export of major histocompatibility class II molecules[127–132]. They are mostly located in the cytoplasm and have crucial roles in major cellular processes: development, energy production, cytoskeletal regulation and immunity. Another interesting cluster involves genes highly expressed in the reproductive tissues: cervix, prostate, uterus and vagina. This cluster is formed by the homeobox A family of genes (*HOXA13*, *HOXA10*, *HOXA11*), *GREB1* and *RPRM*[133–135]. The homeobox A genes are DNA-binding transcription factors important in development, with *HOXA11* being involved in uterine development[134,136]. *RPRM* is a repressor of the mitotic cell cycle[133]. Lastly, *GREB1* is involved in the oestrogen signalling pathway[135]. A third cluster of interest represents genes almost exclusively expressed in brain tissues. This cluster includes repressors of cell proliferation (*RASL10A* and *CGREF1*) and of dendritic growth (*DPYSL5*)[137–139].

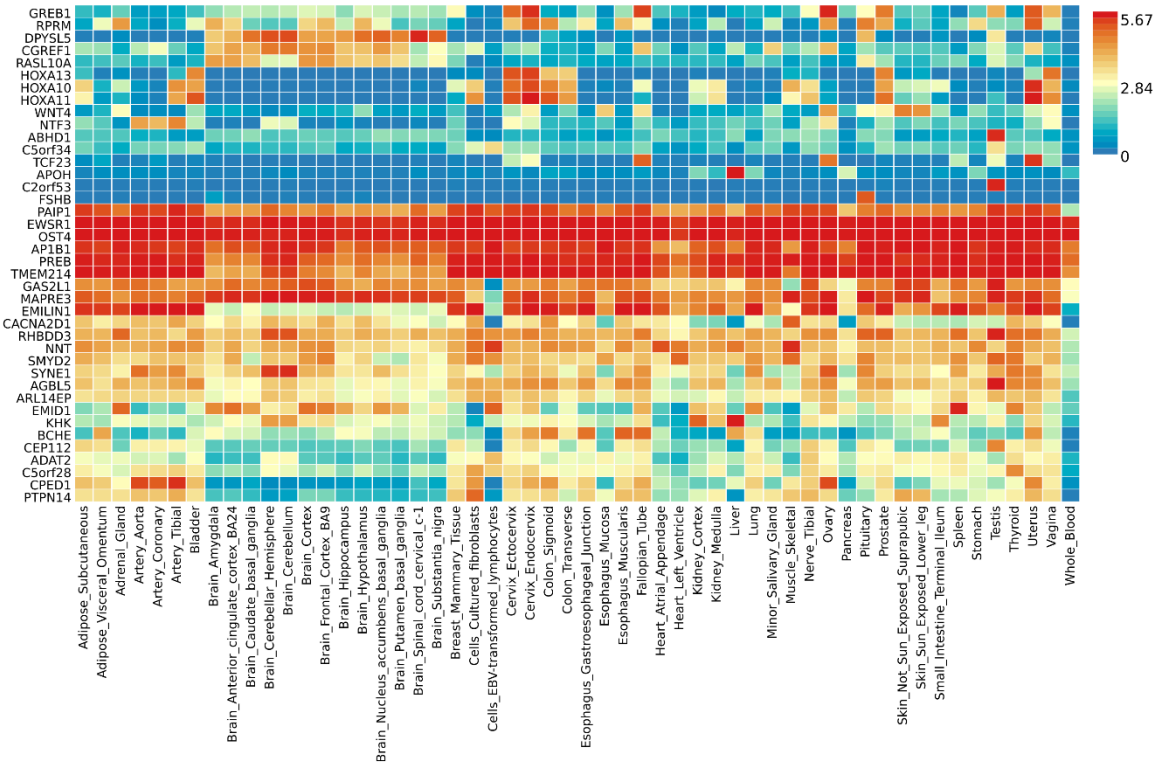


Figure 15. Heatmap of genes expression for genes of interest in 54 tissue types from GTEx v8(56). Lead variants ($p < 10^{-5}$) from multi trait association analysis were mapped positionally to genes by FUMA. Figure was produced from FUMA.

In Figure 16, genes were mapped based on eQTL evidence, which greatly expanded the number of genes being analysed. Similar clusters can be observed as in Figure 15. At the top of the heatmap, homeobox A genes are clustered within reproductive tissues. In the middle, there is the cluster of cells expressed almost exclusively in the brain, *CGREF1* and *DPYSL5*. At the bottom of the heatmap lies the cluster of ubiquitously expressed genes, with important roles in cell survival, metabolism, cytoskeleton structure, transcription and translation, and protein trafficking.

Gene enrichment analysis, utilising data from the GWAS Catalog, identified brain morphology measures as phenotypes that were enriched with 14 target genes that were mapped from nominal SNPs associated with fibromyalgia (Figure 17)[67]. Furthermore, genes were enriched in known gene-sets associated with

gynaecological conditions such as endometriosis and uterine fibroids which also have pain components.

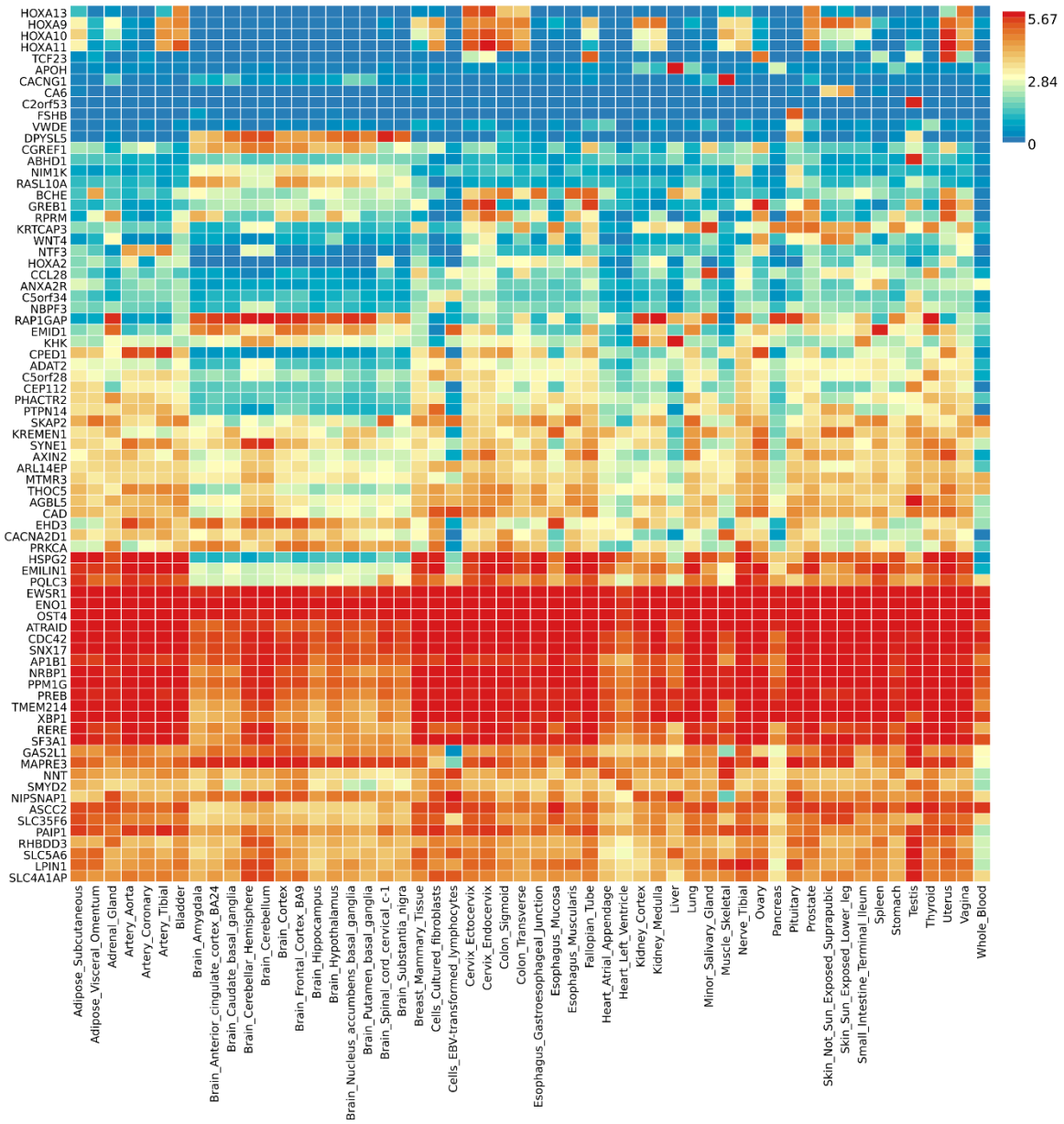


Figure 16. Heatmap of gene expression for genes of interest in 54 tissues from GTEx v8(56). Lead variants ($p < 10^{-5}$) from multi trait association analysis have been mapped based on expression quantitative trait loci from BRAINEAC and eQTLGen databases(100,101). Figure was produced by FUMA.

Top enriched tissues for increased expression of identified target genes were colon (p_{adj} = 0.009) and bladder (p_{adj} = 0.038) (Bonferroni corrected p-value < 0.05) (Figure 18).

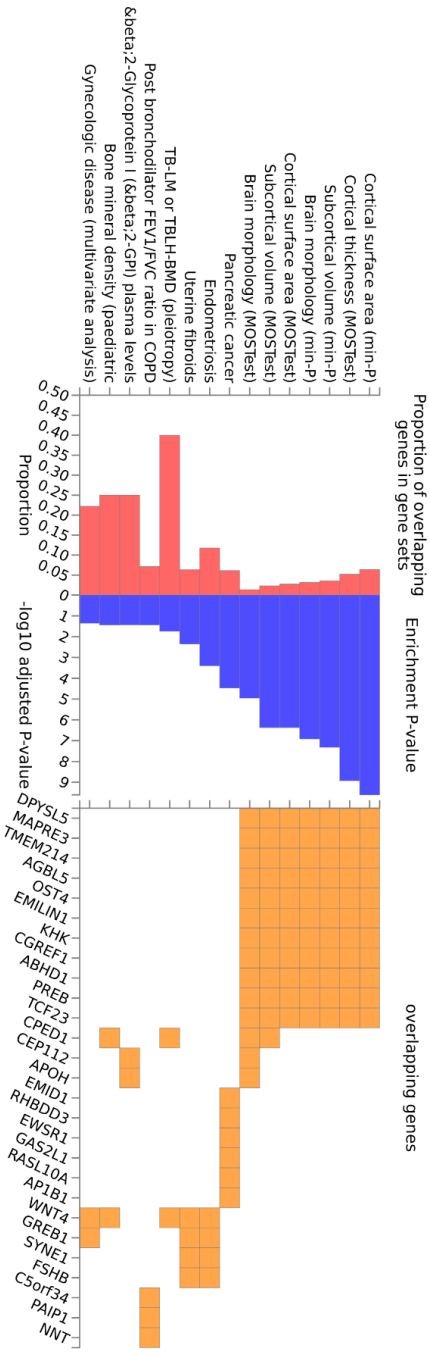


Figure 17. Gene enrichment analysis of GWAS reported genes. The y axis represents gene sets with adjusted P-value < 0.05. Figure was produced by FUMA.

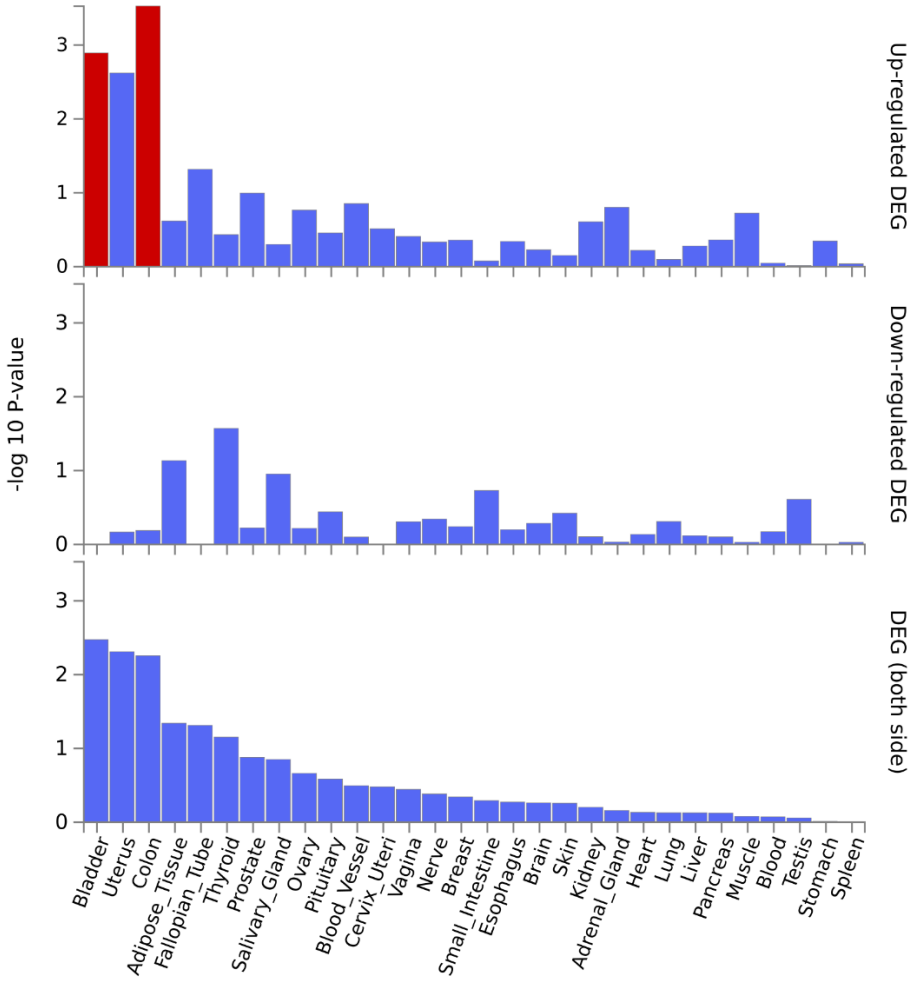


Figure 18. Differentially expressed gene sets analysis. The gene sets are determined by selecting all significant genes, with a Bonferroni corrected p-value < 0.05 and absolute log fold change >= 0.58, from a two-sided t test per tissue versus remaining tissues. Red bars represent significantly enriched tissues with Bonferroni corrected p-values (p < 0.05). Figure was produced by FUMA.

Shared genome-wide significant loci between fibromyalgia and endometriosis

The two genome-wide significant loci for fibromyalgia were looked up in the endometriosis MTAG results which demonstrated that both rs17082358 intronic to *SYNE1* and rs13432756 near *GREB1* (Figure 14) are shared loci between fibromyalgia and endometriosis. The variant, rs17082358, was the top hit in GREB1 locus in both endometriosis and fibromyalgia. The LD between genome-wide

significant lead fibromyalgia variants and genome-wide significant endometriosis variants were 0.024 for rs13432756 and rs11674184 in *SYNE1* locus, suggesting two independent signals in this locus for endometriosis[70].

4.3.4. Biology of shared variants between fibromyalgia and endometriosis

S*YNE1* and *GREB1* are two well-established endometriosis loci. *GREB1* encodes a gene responsible for early response to estrogen signalling in hormone-responsive tissues and in cancer[135]. *SYNE1* encodes a modular protein which plays an important role in subcellular spatial organization by forming a complex with other proteins to connect the nuclear lamina and the cytoskeleton[140]. The heatmap from Figure 14 shows its tissue pattern of expression is ubiquitous, with the highest expression levels recorded in the cerebellum and cerebellar hemisphere. *GREB1* is expressed in low amount in the brain, but highly in reproductive tissues, such as ectocervix, endocervix, the fallopian tubes, ovaries, prostate, testis, uterus and vagina (Figure 13). According to GWAS Catalog, *GREB1* is also associated with several reproductive tract-related traits, including uterine fibroids, ovarian cysts, testosterone levels, uterine and breast cancer, and even vaginal microbiome measurement[67]. Similarly, *SYNE1* is also associated with uterine fibroids and uterine cancer[140].

Interestingly, *GREB1* has been identified as a predictive gene for fibromyalgia identified using differential gene expression between patients with fibromyalgia and healthy controls[141]. S*YNE1* has not been studied in the context of fibromyalgia so far.

4.4. Discussion

4.4.1. Genetic correlation

In this chapter, I have investigated the genetic correlation between fibromyalgia and endometriosis and identified novel genetic variants associated with fibromyalgia and whether they are shared with endometriosis. I have also touched upon the potential shared biology of fibromyalgia and endometriosis. Utilising the largest GWAS meta-analysis of endometriosis to date (Rahmioglu *et al*, 2023) and the GWAS meta-analysis I conducted for fibromyalgia from FinnGen and UK Biobank (see Chapter III), they were identified as highly genetically correlated conditions (rg = 0.46)[40]. This is in line with epidemiological evidence showing that these conditions are highly comorbid[65,117,118,142,143]. Both endometriosis and fibromyalgia are complex conditions with significant genetic heritability contributing to their risk. Although an exact estimate of heritability of fibromyalgia has not been published, a previous twin study has shown a strong additive genetic effect (~50%), while another study has shown heritability estimates vary with age of diagnosis, being highest in people under 40 (28%)[142,143]. Given this, the significant genetic correlation between the two conditions I am highlighting here for the first-time hints to potential shared genetic mechanisms.

4.4.2. Multi trait association analysis

MTAG is useful in boosting power of detection of associations when conditions are highly genetically correlated and has the advantage of running on summary statistics[119]. It makes use of genetically correlated traits to generate trait-specific effect estimates for each SNP. False positives can occur if a SNP is truly null for one trait but non-null for another. To study this, I investigated mean X^2 statistics and maxFDR for both traits and compared p-values and effect sizes between the MTAG

results for fibromyalgia and the meta-analysis of endometriosis and of fibromyalgia, and the univariate GWAS of fibromyalgia from UK Biobank and FinnGen. To identify likely false positives, I compared p-values between MTAG and univariate GWAS for both traits. Variants with non-significant p-values in the univariate GWAS of fibromyalgia, but with very low p-values in the MTAG of fibromyalgia were considered likely false positives. The mean X^2 statistics and maxFDR suggest endometriosis meta-analysis is well powered and fibromyalgia MTAG results likely have inflated effect sizes and are more likely violating the homogeneous-Ω assumption. This suggests MTAG results for fibromyalgia should be interpreted with caution and likely contain false positives. Although maxFDR is a method of probing whether MTAG effect sizes are inflated and if false positives are likely, replication is the best method of assessing whether a SNP is truly associated with the trait studied[119].

There were two genome-wide significant associations with fibromyalgia from MTAG analysis, namely rs17082358 intronic to *SYNE1* on chromosome 6 and rs13432756 intronic to *GREB1* on chromosome 2. rs13432756 is identified as a *cis*-eQTL in blood for *PQLC3* (eQTLGen). S*YNE1* and *GREB1* are both established endometriosis genes with important roles in cytoskeletal structure and estrogen signalling respectively[132,135,144]. Given the strength of correlation between endometriosis and fibromyalgia, uncovering only two genome-wide significant signals is surprising. There are several potential explanations for this finding. Firstly, the meta-analysis of fibromyalgia contains two studies that show some heterogeneity. The resulting meta-analysis showed very few nominally significant hits that could be boosted by MTAG over the threshold of genome-wide significance. Secondly, the meta-analysis of fibromyalgia is underpowered compared to the meta-

analysis of endometriosis. Thirdly, the SNP-heritability of fibromyalgia may be overestimated, which can sometimes lead to a very high genetic correlation.

4.4.3. Biology of variants of interest

The two genome-wide significant hits are interesting in terms of biology and potential link to fibromyalgia. Hormones, especially estrogen, have long been investigated for their role in pain modulation and transition from acute to chronic pain. Estrogen's role in pain mechanisms appears to be complex and can either ameliorate pain or lead to hyperalgesia[145–147]. In one study of dysmenorrhea, hormonal contraception was found to be slightly protective of comorbidity with fibromyalgia, illustrating a link between fibromyalgia and hormone regulation[148]. However, in a double-blind study of 17β-oestradiol aimed at reducing pain in postmenopausal women with fibromyalgia, they showed no differences between control and treatment group in terms of self-reported pain, quantitative sensory testing, or cold pressor test results[149]. Hormonal contraception contains progesterone as well as estrogen and according to a study by Schertzinger *et al.* (2018), they showed low progesterone levels were associated with increased self-reported pain[150]. Interestingly, women reported least pain in the pressure point test during the luteal phase of the menstrual cycle, where progesterone is at its highest, and most pain sensitive in the early follicular phase, when progesterone and estrogen are lowest[151].

Next, FUMA identified 22 nominally associated ($p < 10^{-5}$) fibromyalgia lead SNPs, which were mapped to genes either positionally or according to eQTL evidence from eQTLGen or BRAINEAC[17,73,102]. These mapped genes were classified based on tissue expression from GTEx v8 into four major clusters: ubiquitous, brain-specific, reproductive organ-specific or varied levels of expression throughout visceral tissues[75]. These genes highlight the potential biology relative to fibromyalgia and

have varied, integral roles in the cell: from mediating cell survival to metabolism. Given they are common variants that affect likely important, non-redundant genes, their effect sizes were small (beta < 0.1), pointing to the low burden, common variant hypothesis in complex diseases. Moreover, their effects are likely pleiotropic between fibromyalgia and endometriosis, with the p-values of the SNPs showing at least nominal genome-wide significance in both the meta-analysis of endometriosis and fibromyalgia.

Although many lead SNPs are eQTLs in the blood and brain, a recent study has shown that there are important differences between GWAS hits and eQTLs[152]. The authors have shown GWAS hits are near genes with key functional annotations, such as transcription factors, and are under strong selective constraint. On the other hand, eQTLs tend to localise near promoters of genes without key functions. Moreover, the impact of a variant on a gene may be limited to specific conditions, such as developmental stage, environmental triggers, or in specific cell types. Inferring gene impact from eQTLs is a popular method, however some GWAS hits are not eQTLs. This is the case with one of the genome-wide significant variants from fibromyalgia MTAG results (rs17082358 from *SYNE1* locus). This may also be explained by eQTL maps being underpowered to detect associations in disease-relevant tissues to fibromyalgia. Alternatively, the positional evidence we have illustrates *SYNE1* or some other near-by genes (e.g. *ESR1*) as the likely target genes which are highly expressed in neuronal and neuroendocrine cells, immune cells and epithelial cells. Unlike in the previous study, the lead SNPs are eQTLs which impact many genes with key functional roles.

Differential expression of gene sets analysis showed bowel and bladder as being enriched in the genes of interest. A potential explanation for this is many patients

with either endometriosis or fibromyalgia have other chronic pain conditions, such as irritable bowel syndrome or bladder pain syndrome[62]. Therefore, it is possible these genes are part of a common network of pain genes that affect many different tissue types and result in comorbidities.

4.4.4. Conclusions

In this chapter, the genetic correlation and shared genetic basis of fibromyalgia and endometriosis was explored. Fibromyalgia and endometriosis were shown to be strongly genetically correlated, as predicted by their high comorbidity, but given limitations from our meta-analysis, this may be an artefact. Compared to the previous chapter, where a meta-analysis of fibromyalgia has shown no genome-wide significant hits, here I identified two genome-wide significant SNPs, within *SYNE1* and *GREB1*, respectively, which have been associated previously with endometriosis. I identified 22 lead SNPs, which were eQTLs in both blood and brain. The mapped genes are widely expressed in many tissues, and many have key functional roles.

Chapter V. Discussion and Conclusions

5.1. Discussion

5.1.1. Key findings

The aim of my thesis is to study the genetic basis of fibromyalgia, generally considered a primary pain condition, and its potential shared genetic basis with, endometriosis, a secondary pain condition. COPCs are a group of primary and secondary pain conditions recognised as important research topics due to their high rate of comorbidity and prevalence in society, which includes fibromyalgia and endometriosis[61,62,153]. These two conditions have not been extensively studied

together, are often comorbid, and have a strong heritable component (see Section 1.1)[25,50,65,66]. Moreover, fibromyalgia has not had a large GWAS performed previously, therefore making it an interesting candidate for analysis.

In the second chapter, I conducted a systematic review of the literature on primary pain conditions GWAS. There were several variants with likely pleiotropic effects in different primary pain conditions: dysmenorrhea, multisite chronic pain, migraine, and chronic headaches. The most likely candidate for future studies is rs7523086, intronic to *NGF* gene. It is associated with endometriosis, migraine, and dysmenorrhea and NGF was shown to play a role in inflammation and endometriosis associated pain. Its role in the adult central nervous system is an open question, however it is important in the development of sensory neurons. Other pain vulnerability genes discussed were *ASTN2* and *LRP1*. The latter is known to be involved in neuropathic pain and mediates an inflammatory environment in endometriosis. By contrast, the role of ASTN2 in chronic pain is not presently known and therefore, requires further investigation. This systematic review of primary pain conditions GWAS had therefore linked variant rs7523086 and genes *NGF*, *ASTN2* and *LRP1* to multiple pain conditions, strengthening the evidence of association with pain, rather than another pathophysiological mechanism of the conditions.

GWAS reveal a wealth of information about the genetic landscape of chronic pain conditions. However, in the case of fibromyalgia, the only published GWAS identified was underpowered and did not find any genome-wide associations. Therefore, in Chapter III, I investigated the genetics of fibromyalgia in UK Biobank, investigating in depth the results from different case definitions and their differences in genetic associations between definitions, and conducted a meta-analysis with publicly available summary statistics of fibromyalgia from FinnGen.

I first aggregated the case definitions of fibromyalgia in order to investigate sex-specific signals and identified different male and female nominal genetic associations with fibromyalgia. Sensitivity analyses revealed genetic heterogeneity between clinical records of fibromyalgia from primary and secondary care, and self-reported cohorts. These differences were further investigated by looking at different population characteristics, such as age, female:male ratio, BMI, ethnicity, a subjective rating of overall health or educational attainment. The ICD cohort reported significantly worse overall health than primary care cohort, prompting the hypothesis that secondary care cases included more severe disease.

The true cause of genetic heterogeneity between different case definitions is not known. Alternative hypotheses for this discovery are that these are false positives uncovered by chance; the case definitions represent different musculoskeletal conditions; or they are true fibromyalgia cases, but at different points on a fibromyalgia severity scale.

Different methods of conducting GWAS were also compared as part of my thesis. In summary, SAIGE and REGENIE tools showed consistent and less inflated effect sizes than BOLT-LMM due to their underlying algorithms. REGENIE was computationally faster and easier to use and was the GWAS method of choice for the present analyses.

Meta-analysis of fibromyalgia GWAS was performed with the ICD-code identified cases from UK Biobank and FinnGen cohorts. One nominal association was discovered on chromosome 7, rs34323745, located in *GPNMB* locus ($p = 7.269 \times 10^{-8}$).

In Chapter IV, I showed that endometriosis and fibromyalgia are genetically correlated. Unfortunately, the boost in power conferred by MTAG with endometriosis did not add more evidence of association for the top SNP from the meta-analysis. Two genome-wide hits were identified with MTAG for fibromyalgia, rs17082358, an intronic SNP within *SYNE1* gene on chromosome 6 ($p = 1.49 \times 10^{-11}$) and rs13432756, an intronic SNP within *GREB1* gene on chromosome 2 ($p = 6.39 \times 10^{-9}$). Only GREB1 was previously associated with fibromyalgia, although the mechanism of this finding was not explored, neither was this finding replicated before[141]. These variants, together with the nominally associated variants, point to a shared network of genes involved in the pathologies of endometriosis and fibromyalgia involving hormonal regulation. They map to genes which show either ubiquitous or tissue-specific expression. Ubiquitously expressed genes also have key roles in the cell, such as in maintaining the structure of the cytoskeleton, vesicle transport, transcription, and translation, etc. Tissue-specific clusters of genes form around increased brain or reproductive organs expression.

In Chapter I, I identified *NGF*, *ASTN2* and *LRP1* as potential pain susceptibility genes, with important roles in peripheral and central nervous system with *NGF* being associated with migraines and dysmenorrhea. In Chapter IV, I could not identify any variants in these genes that are associated with fibromyalgia from the MTAG results. However, I identified *SYNE1* and *GREB1* as shared loci associated with fibromyalgia and endometriosis. Both *NGF* and *SYNE1* have been shown to contribute to a pro-inflammatory environment in the pelvis. Some ovarian cancer patients have a higher burden of mutations in this gene which may contribute to a pro-inflammatory response[140]. Similarly, *NGF* signalling can promote inflammation in endometriosis-associated pain[59]. G*REB1* codes an early response protein in estrogen signalling

and is potentially associated with fibromyalgia, a finding which needs to be replicated[141]. In summary, *NGF*, *GREB1* and *SYNE1* are potentially important in peripheral physiology of endometriosis, and the latter two, in fibromyalgia as well. S*YNE1* is also highly expressed in the brain, suggesting a central mechanism of pain modulation as well as peripheral.

5.1.2. Limitations

One limitation of this study is that when we talk about endometriosis, there is no information on the pain status of the patients. Therefore, when we are analysing fibromyalgia and endometriosis together, we are not distinguishing between those with pain and those without, therefore missing potential important differences. It is important for genetic discovery of genes underlying pain susceptibility to categorise patients based on pain status.

Limitations of working with UK Biobank data are the unknown completeness and coverage of the primary care and hospital in-patient data. Fibromyalgia numbers are therefore likely an underestimate. Prevalence of fibromyalgia in UK Biobank is around 1%, whereas in the general population it is estimated to be in the range of 2-6%. Males are specifically underdiagnosed with fibromyalgia and true male:female fibromyalgia ratios are closer to 1:1[47].

Another limitation is the heterogeneity between UK Biobank and FinnGen GWAS, even when limiting the analysis to the ICD cohort. The source of the heterogeneity is not known, and the meta-analysis results should be interpreted with caution.

A limitation of MTAG analysis is that it assumes that non-null SNPs for one trait are also not null for the other trait. SNPs where this assumption is violated could be false positives. Indeed, I identified a few potential false positive associations, which were

nominal MTAG SNPs where the p-value of association is very low in endometriosis, but non-significant in fibromyalgia. The best method to confirm these findings is through replication studies, ideally in a bigger cohort.

5.1.3. Future work

The pain questionnaire in UK Biobank contains questions from the ACR 1990. This is another source of potential fibromyalgia cases in UK Biobank, with the caveat that the questions pertain to symptoms in the last three months, and there is no clinical evidence on a case-by-case basis that could corroborate the results. However, due to time constraints, I focused solely on direct evidence of fibromyalgia. This was obtained either by asking if a doctor has told the participant they have a long-term illness, if they've ever been diagnosed with this condition, or whether there is a medical record of a diagnosis. An important conclusion of this work is that fibromyalgia in UK Biobank is highly heterogenous between the different diagnostic groups and when compared to other fibromyalgia GWAS. If future studies will use fibromyalgia data from UK Biobank, this caveat should be considered, and any conclusions should be interpreted with caution.

A hypothesis for the observed genetic heterogeneity is that either hospital or primary care cases are more severe. This could be further assessed by studying the overlap between the different case definitions and other chronic pain conditions. A higher burden of chronic pain conditions could be indicative of more severe cases.

Primary care data is an important source of cases in UK Biobank. Although there are genetic differences between primary care and secondary care cases, they should not be discounted as they likely represent real cases. An interesting point of comparison would be looking at FinnGen hospital versus primary care genetic heterogeneity. The

population size is comparable and the access to patient records is more comprehensive than in UK Biobank. Other biobanks, such as Estonian Biobank, would also serve as a useful comparison and future work could look at a meta-analysis between Estonian Biobank, FinnGen and UK Biobank.

What is evident from Chapters III and IV is the need for larger fibromyalgia cohorts with better phenotype information that can be used to design well powered GWAS. Current cohort sizes in both UK Biobank and FinnGen are small, and this has impacted the power of discovery of genetic associations. What this study has shown is that fibromyalgia may be a more heterogeneous condition than thought and this has had an impact on the GWAS and meta-analysis conducted here.

Future analysis could also focus on replicating and putting the results of MTAG into a biological context. This could be accomplished with different methods. It was previously shown that GWAS hits often do not overlap eQTLs, therefore, different annotation resources could involve studying overlap with methylation sites or known enhancers. More experiments that cover different developmental stages, cell types and different environmental factors could improve detection of eQTLs. In my analysis, many nominal associations are eQTLs in a plethora of tissues relevant to endometriosis and fibromyalgia, such as brain, skeletal muscle, blood and reproductive organs, however, this not the case for one of the top GWAS hits. Another idea is to investigate whether the GWAS hits from MTAG are also genome-wide significant in the image-derived phenotypes GWAS from 40,000 participants of UK Biobank[154]. Focusing on areas with gray matter alterations would narrow the focus of the search. These have been studied extensively and examples include bilateral medial frontal gyri, bilateral superior frontal gyri, right pre- and post-central gyri (including the S1 and primary motor cortex), bilateral insula (anterior), right

cingulate cortex (dorsal posterior cingulate cortex), basal ganglia, thalamus, and periaqueductal gray[15]. Next, Mendelian Randomisation could be utilized in patients with overlapping endometriosis and fibromyalgia to determine whether variants associated with either one is causal to the other condition[39].

In addition to endometriosis and fibromyalgia, there are other comorbid chronic pain conditions which could share genetic risk factors. Some COPCs are well represented in UK Biobank, such as irritable bowel syndrome, migraine, low back pain, chronic tension type headaches and temporomandibular joint disorder. Other comorbidities of interest are bladder pain syndrome and vulvodynia, however case numbers in UK Biobank are too small for a sufficiently powered GWAS (<500 cases). Other cohorts should be considered for genetic analysis of these conditions. Multi-trait analysis of all COPCs would be a great source of discovery of common genes involved in mechanisms of nociplastic pain. This analysis could also reveal whether genetic heterogeneity between different case definitions applies to other chronic pain conditions.

5.2. Conclusions

There is genetic overlap between endometriosis and fibromyalgia, in the *SYNE1* and *GREB1* loci. These conditions are often comorbid and patients with both could benefit from pain treatments that consider all the dimensions of their pain. Lastly, fibromyalgia in UK Biobank was shown here to be highly heterogeneous and this should be taken into account in further studies.

References

1. Public Health England. Chronic pain in adults. (2017).

2. Denk, F., McMahon, S. B. & Tracey, I. Pain vulnerability: a neurobiological perspective. *Nat Neurosci* 17, 192–200 (2014).

3. Burri, A., Ogata, S., Rice, D. & Williams, F. M. K. Twelve-year follow-up of chronic pain in twins: Changes in environmental and genetic influence over time. *European Journal of Pain* 22, 1439–1447 (2018).

4. Junqueira, D. R. G. *et al.* Heritability and lifestyle factors in chronic low back pain: Results of the Australian Twin Low Back Pain Study (The AUTBACK study). *European Journal of Pain* 18, 1410–1418 (2014).

5. Fagnani, C. *et al.* An Italian Twin Study of Non-Cancer Chronic Pain as a Wide Phenotype and Its Intensity. *Medicina* 58, 1522 (2022).

6. Nicholas, M. *et al.* The IASP classification of chronic pain for ICD-11: chronic primary pain. *Pain* 160, (2019).

7. Treede, R.-D. *et al.* Chronic pain as a symptom or a disease: the IASP Classification of Chronic Pain for the International Classification of Diseases (ICD-11). *Pain* 160, 19–27 (2019).

8. Lee, M. & Tracey, I. Neuro-genetics of persistent pain. *Curr Opin Neurobiol* 23, 127–132 (2013).

9. Reichling, D. B. & Levine, J. D. Critical role of nociceptor plasticity in chronic pain. *Trends Neurosci* 32, 611–618 (2009).

10. Tracey, I. Can neuroimaging studies identify pain endophenotypes in humans? *Nat Rev Neurol* 7, 173–181 (2011).

11. Basbaum, A. I., Bautista, D. M., Scherrer, G. & Julius, D. Cellular and Molecular Mechanisms of Pain. *Cell* 139, 267–284 (2009).

12. Harte, S. E., Harris, R. E. & Clauw, D. J. The neurobiology of central sensitization. *Journal of Applied Biobehavioral Research* 23, e12137 (2018).

13. Treede, R.-D., Hoheisel, U., Wang, D. & Magerl, W. Central sensitization: clinical utility of a physiological concept for the International Statistical Classification of Diseases and Related Health Problems and for nociplastic pain. *PAIN* 163, S99 (2022).

14. Cohen, S. P., Vase, L. & Hooten, W. M. Chronic pain: an update on burden, best practices, and new advances. *Lancet* 397, 2082–2097 (2021).

15. Cauda, F. *et al.* Gray matter alterations in chronic pain: A network-oriented meta-analytic approach. *Neuroimage Clin* 4, 676–686 (2014).

16. Gracely, R. H. *et al.* Pain catastrophizing and neural responses to pain among persons with fibromyalgia. *Brain* 127, 835–843 (2004).

17. Brandl, F. *et al.* Common and specific large-scale brain changes in major depressive disorder, anxiety disorders, and chronic pain: a transdiagnostic multimodal meta-analysis of structural and functional MRI studies. *Neuropsychopharmacology* 47, 1071–1080 (2022).

18. Zheng, C. J., Drunen, S. V. & Egorova-Brumley, N. Neural correlates of co-occurring pain and depression: an activation-likelihood estimation (ALE) meta-analysis and systematic review. *Translational Psychiatry* 12, 1–16 (2022).

19. Zunhammer, M., Spisák, T., Wager, T. D. & Bingel, U. Meta-analysis of neural systems underlying placebo analgesia from individual participant fMRI data. *Nature Communications* 12, 1–11 (2021).

20. Amanzio, M. & Palermo, S. Pain Anticipation and Nocebo-Related Responses: A Descriptive Mini-Review of Functional Neuroimaging Studies in Normal Subjects and Precious Hints on Pain Processing in the Context of Neurodegenerative Disorders. *Frontiers in Pharmacology* 10, (2019).

21. Bee, L. A. & Dickenson, A. H. Descending facilitation from the brainstem determines behavioural and neuronal hypersensitivity following nerve injury and efficacy of pregabalin. *Pain* 140, 209–223 (2008).

22. Gardell, L. R. *et al.* Enhanced evoked excitatory transmitter release in experimental neuropathy requires descending facilitation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 23, 8370–8379 (2003).

23. Burgess, S. E. *et al.* Time-dependent descending facilitation from the rostral ventromedial medulla maintains, but does not initiate, neuropathic pain. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 22, 5129–5136 (2002).

24. Tracey, I. Imaging pain. *BJA: British Journal of Anaesthesia* 101, 32–39 (2008).

25. Saha, R. *et al.* Heritability of endometriosis. *Fertility and Sterility* 104, 947–952 (2015).

26. McIntosh, A. M. *et al.* Genetic and Environmental Risk for Chronic Pain and the Contribution of Risk Variants for Major Depressive Disorder: A Family-Based Mixed-Model Analysis. *PLOS Medicine* 13, e1002090 (2016).

27. Hocking, L. J., Morris, A. D., Dominiczak, A. F., Porteous, D. J. & Smith, B. H. Heritability of chronic pain in 2195 extended families. *European Journal of Pain (London, England)* 16, 1053–1063 (2012).

28. Mogil, J. S. Pain genetics: past, present and future. *Trends Genet* 28, 258–266 (2012).

29. March, R. Gene mapping by linkage and association analysis. *Molecular Biotechnology* 13, 113–122 (1999).

30. Mitchell, K. J. What is complex about complex disorders? *Genome Biology* 13, (2012).

31. Ioannidis, J. P. A., Trikalinos, T. A. & Khoury, M. J. Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *American Journal of Epidemiology* 164, 609–614 (2006).

32. Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S. & Hirschhorn, J. N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genetics* 33, 177–182 (2003).

33. Daly, A. K. & Day, C. P. Candidate gene case-control association studies: advantages and potential pitfalls. *British journal of clinical pharmacology* 52, 489–499 (2001).

34. Anderson, C. A. *et al.* Data quality control in genetic case-control association studies. *Nature protocols* 5, 1564–1573 (2010).

35. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics* 101, 5–22 (2017).

36. McCarthy, M. I. & Hirschhorn, J. N. Genome-wide association studies: past, present and future. *Hum Mol Genet* 17, R100-101 (2008).

37. Zorina-Lichtenwalter, K., Meloto, C. B., Khoury, S. & Diatchenko, L. Genetic predictors of human chronic pain conditions. *Neuroscience* 338, 36–62 (2016).

38. Farrell, S. F. *et al.* A Shared Genetic Signature for Common Chronic Pain Conditions and its Impact on Biopsychosocial Traits. *J Pain* S1526-5900(22)00431-X (2022) doi:10.1016/j.jpain.2022.10.005.

39. Thanassoulis, G. & O'Donnell, C. J. Mendelian Randomization: Nature's Randomized Trial in the Post–Genome Era. *JAMA* 301, 2386–2388 (2009).

40. Rahmioglu, N. *et al.* The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions. *Nature genetics* 55, 423–436 (2023).

41. Clauw, D. J., Arnold, L. M. & McCarberg, B. H. The Science of Fibromyalgia. *Mayo Clinic Proceedings* 86, 907–911 (2011).

42. Russell, I. J. & Larson, A. A. Neurophysiopathogenesis of fibromyalgia syndrome: a unified hypothesis. *Rheumatic Diseases Clinics of North America* 35, 421–435 (2009).

43. Qureshi, A. G. *et al.* Diagnostic Challenges and Management of Fibromyalgia. *Cureus* 13, e18692.

44. Wolfe, F. *et al.* Fibromyalgia criteria and severity scales for clinical and epidemiological studies: a modification of the ACR Preliminary Diagnostic Criteria for Fibromyalgia. *The Journal of Rheumatology* 38, 1113–1122 (2011).

45. Wolfe, F. *et al.* The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. *Arthritis Care & Research* 62, 600–610 (2010).

46. Marques, A. P., Santo, A. de S. do E., Berssaneti, A. A., Matsutani, L. A. & Yuan, S. L. K. Prevalence of fibromyalgia: literature review update. *Rev Bras Reumatol Engl Ed* 57, 356–363 (2017).

47. Wolfe, F., Walitt, B., Perrot, S., Rasker, J. J. & Häuser, W. Fibromyalgia diagnosis and biased assessment: Sex, prevalence and bias. *PloS one* 13, e0203755 (2018).

48. Gyorfi, M., Rupp, A. & Abd-Elsayed, A. Fibromyalgia Pathophysiology. *Biomedicines* 10, 3070 (2022).

49. Sarzi-Puttini, P., Giorgi, V., Marotto, D. & Atzeni, F. Fibromyalgia: an update on clinical characteristics, aetiopathogenesis and treatment. *Nature Reviews Rheumatology* 16, 645–660 (2020).

50. Dizner-Golab, A., Lisowska, B. & Kosson, D. Fibromyalgia - etiology, diagnosis and treatment including perioperative management in patients with fibromyalgia. *Reumatologia* 61, 137–148 (2023).

51. D'Agnelli, S. *et al.* Fibromyalgia: Genetics and epigenetics insights may provide the basis for the development of diagnostic biomarkers. *Molecular Pain* 15, 1744806918819944 (2018).

52. Knezevic, N. N., Tverdohleb, T., Knezevic, I. & Candido, K. D. The Role of Genetic Polymorphisms in Chronic Pain Patients. *International Journal of Molecular Sciences* 19, 1707 (2018).

53. Docampo, E. *et al.* Genome-wide analysis of single nucleotide polymorphisms and copy number variants in fibromyalgia suggest a role for the central nervous system. *PAIN* 155, 1102–1109 (2014).

54. Arnold, L. M. *et al.* The Fibromyalgia Family Study: A Genome-Scan Linkage Study. *Arthritis and rheumatism* 65, 1122–1128 (2013).

55. Smith, S. B. *et al.* Large candidate gene association study reveals genetic risk factors and therapeutic targets for fibromyalgia. *Arthritis and Rheumatism* 64, 584–593 (2012).

56. Rutigliano, G., Accorroni, A. & Zucchi, R. The Case for TAAR1 as a Modulator of Central Nervous System Function. *Frontiers in Pharmacology* 8, 987 (2017).

57. Zhang, Y. *et al.* Trace amine-associated receptor 1 regulation of Kv1.4 channels in trigeminal ganglion neurons contributes to nociceptive behaviors. *The Journal of Headache and Pain* 24, 49 (2023).

58. Zondervan, K. T. *et al.* Endometriosis. *Nature Reviews Disease Primers* 4, 1–25 (2018).

59. Maddern, J., Grundy, L., Castro, J. & Brierley, S. M. Pain in Endometriosis. *Frontiers in Cellular Neuroscience* 14, 590823 (2020).

60. Vercellini, P. *et al.* Association between endometriosis stage, lesion type, patient characteristics and severity of pelvic pain symptoms: a multivariate analysis of over 1000 patients. *Human Reproduction (Oxford, England)* 22, 266–271 (2007).

61. Institute of Medicine (US) Committee on Advancing Pain Research, Care, and Education. *Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research*. (National Academies Press (US), Washington (DC), 2011).

62. Maixner, W., Fillingim, R. B., Williams, D. A., Smith, S. B. & Slade, G. D. Overlapping Chronic Pain Conditions: Implications for Diagnosis and Classification. *J Pain* 17, T93–T107 (2016).

63. Kato, K., Sullivan, P. F., Evengård, B. & Pedersen, N. L. A population-based twin study of functional somatic syndromes. *Psychological Medicine* 39, 497–505 (2009).

64. Diatchenko, L., Nackley, A. G., Slade, G. D., Fillingim, R. B. & Maixner, W. Idiopathic pain disorders--pathways of vulnerability. *Pain* 123, 226–230 (2006).

65. Greenbaum, H., Weil, C., Chodick, G., Shalev, V. & Eisenberg, V. H. Evidence for an association between endometriosis, fibromyalgia, and autoimmune diseases. *American Journal of Reproductive Immunology (New York, N.Y.: 1989)* 81, e13095 (2019).

66. Coloma, J. L. *et al.* Prevalence of fibromyalgia among women with deep infiltrating endometriosis. *International Journal of Gynecology & Obstetrics* 146, 157–163 (2019).

67. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 47, D1005–D1012 (2019).

68. Little, J. *et al.* STrengthening the REporting of Genetic Association Studies (STREGA)--an extension of the STROBE statement. *Genet Epidemiol* 33, 581–598 (2009).

69. Auton, A. *et al.* A global reference for human genetic variation. *Nature* 526, 68–74 (2015).

70. Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 31, 3555–3557 (2015).

71. Staley, J. R. *et al.* PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* 32, 3207–3209 (2016).

72. Kamat, M. A. *et al.* PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics (Oxford, England)* 35, 4851–4853 (2019).

73. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 8, 1826 (2017).

74. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria).

75. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348, 648–660 (2015).

76. Meloto, C. B. *et al.* Human pain genetics database: a resource dedicated to human pain genetics research. *Pain* 159, 749–763 (2018).

77. Mortlock, S. *et al.* Tissue specific regulation of transcription in endometrium and association with disease. *Hum Reprod* 35, 377–393 (2020).

78. Fung, J. N. *et al.* Genetic regulation of disease risk and endometrial gene expression highlights potential target genes for endometriosis and polycystic ovarian syndrome. *Scientific Reports* 8, 11424 (2018).

79. van Dongen, R. M. *et al.* Migraine biomarkers in cerebrospinal fluid: A systematic review and meta-analysis. *Cephalalgia* 37, 49–63 (2017).

80. Coxon, L., Horne, A. W. & Vincent, K. Pathophysiology of endometriosis-associated pain: A review of pelvic and central nervous system mechanisms. *Best Practice & Research Clinical Obstetrics & Gynaecology* 51, 53–67 (2018).

81. Goadsby, P. J. Pathophysiology of migraine. *Ann Indian Acad Neurol* 15, S15–S22 (2012).

82. Gaultier, A. *et al.* A shed form of LDL receptor–related protein–1 regulates peripheral nerve injury and neuropathic pain in rodents. *The Journal of Clinical Investigation* 118, 161–172 (2008).

83. Bora, G. & Yaba, A. The role of mitogen-activated protein kinase signaling pathway in endometriosis. *The Journal of Obstetrics and Gynaecology Research* 47, 1610–1623 (2021).

84. Chuang, T.-Y. *et al.* LRP1 expression in microglia is protective during CNS autoimmunity. *Acta Neuropathologica Communications* 4, 68 (2016).

85. Inoue, R. *et al.* Effects of rs958804 and rs7858836 single-nucleotide polymorphisms of the ASTN2 gene on pain-related phenotypes in patients who underwent laparoscopic colectomy and mandibular sagittal split ramus osteotomy. *Neuropsychopharmacology Reports* 41, 82–90 (2021).

86. Bourgeron, T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nature Reviews Neuroscience* vol. 16 551 (2015).

87. Rahman, M. S. *et al.* Genome-wide association study identifies RNF123 locus as associated with chronic widespread musculoskeletal pain. *Ann Rheum Dis* annrheumdis-2020-219624 (2021) doi:10.1136/annrheumdis-2020-219624.

88. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 12, e1001779 (2015).

89. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* 613, 508–518 (2023).

90. Bycroft, C. *et al. Genome-Wide Genetic Data on ~500,000 UK Biobank Participants*. http://biorxiv.org/lookup/doi/10.1101/166298 (2017) doi:10.1101/166298.

91. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* 81, 559–575 (2007).

92. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4, s13742-8 (2015).

93. PLINK: Whole genome data analysis toolset. https://zzz.bwh.harvard.edu/plink/contact.shtml.

94. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics* 47, 284–290 (2015).

95. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics* 53, 1097–1103 (2021).

96. Rapid GWAS of thousands of phenotypes for 337,000 samples in the UK Biobank. *Neale lab* http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank.

97. DeBoever, C. *et al.* Assessing Digital Phenotyping to Enhance Genetic Studies of Human Diseases. *Am J Hum Genet* 106, 611–622 (2020).

98. Willer, C., Li, Y. & Abecasis, G. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190–2191 (2010).

99. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 47, 291–295 (2015).

100. Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58 (2010).

101. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nature Genetics* 53, 1300–1310 (2021).

102. Ramasamy, A. *et al.* Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nature neuroscience* 17, 1418–1428 (2014).

103. Li, H. & Burstein, E. COMMD1 regulates inflammation and colitis-associated cancer progression. *OncoImmunology* 3, (2014).

104. Weiskirchen, R. & Penning, L. C. COMMD1, a multi-potent intracellular protein involved in copper homeostasis, protein trafficking, inflammation, and cancer. *Journal of Trace Elements in Medicine and Biology* 65, (2021).

105. Li, H. *et al.* Copper Metabolism Domain-Containing 1 Represses Genes That Promote Inflammation and Protects Mice From Colitis and Colitis-Associated Cancer. *Gastroenterology* 147, 184-195.e3 (2014).

106. Wolfe, F. Fibromyalgianess. *Arthritis Care & Research* 61, 715–716 (2009).

107. Saade, M., Souza, G. A. de, Scavone, C. & Kinoshita, P. F. The Role of GPNMB in Inflammation. *Frontiers in Immunology* 12, 674739 (2021).

108. Ripoll, V. M., Irvine, K. M., Ravasi, T., Sweet, M. J. & Hume, D. A. Gpnmb is induced in macrophages by IFN-gamma and lipopolysaccharide and acts as a feedback regulator of proinflammatory responses. *Journal of Immunology (Baltimore, Md.: 1950)* 178, 6557–6566 (2007).

109. *Endometriosis: Diagnosis and Management*. https://www.nice.org.uk/guidance/ng73 (2017).

110. Sutton, C. J., Ewen, S. P., Whitelaw, N. & Haines, P. Prospective, randomized, double-blind, controlled trial of laser laparoscopy in the treatment of pelvic pain associated with minimal, mild, and moderate endometriosis. *Fertility and Sterility* 62, 696–700 (1994).

111. Laux-Biehlmann, A., d'Hooghe, T. & Zollner, T. M. Menstruation pulls the trigger for inflammation and pain in endometriosis. *Trends in Pharmacological Sciences* 36, 270–276 (2015).

112. Barcz, E., Kamiński, P. & Marianowski, L. Role of cytokines in pathogenesis of endometriosis. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research* 6, 1042–1046 (2000).

113. Brierley, S. M. & Linden, D. R. Neuroplasticity and dysfunction after gastrointestinal inflammation. *Nature Reviews Gastroenterology & Hepatology* 11, 611–627 (2014).

114. Tokushige, N., Markham, R., Russell, P. & Fraser, I. S. Nerve fibres in peritoneal endometriosis. *Human Reproduction (Oxford, England)* 21, 3001–3007 (2006).

115. Ge, P. *et al.* Linaclotide treatment reduces endometriosis-associated vaginal hyperalgesia and mechanical allodynia through viscerovisceral cross-talk. *Pain* 160, 2566–2579 (2019).

116. Petzke, F., Clauw, D. J., Ambrose, K., Khine, A. & Gracely, R. H. Increased pain sensitivity in fibromyalgia: effects of stimulus type and mode of presentation. *Pain* 105, 403–413 (2003).

117. Pardo, F. L., Bondesson, E., Schelin, M. E. C. & Jöud, A. A diagnosis of rheumatoid arthritis, endometriosis or IBD is associated with later onset of fibromyalgia and chronic widespread pain. *European Journal of Pain* 23, 1563–1573 (2019).

118. Sinaii, N., Cleary, S. D., Ballweg, M. L., Nieman, L. K. & Stratton, P. High rates of autoimmune and endocrine disorders, fibromyalgia, chronic fatigue syndrome and atopic

diseases among women with endometriosis: a survey analysis. *Human Reproduction* 17, 2715–2724 (2002).

119. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* 50, 229–237 (2018).

120. Su, L. K. & Qi, Y. Characterization of human MAPRE genes and their proteins. *Genomics* 71, 142–149 (2001).

121. Stroud, M. J. *et al.* GAS2-like proteins mediate communication between microtubules and actin through interactions with end-binding proteins. *Journal of Cell Science* 127, 2672–2682 (2014).

122. Li, C. *et al.* Transmembrane Protein 214 (TMEM214) mediates endoplasmic reticulum stress-induced caspase 4 enzyme activation and apoptosis. *The Journal of Biological Chemistry* 288, 17908–17917 (2013).

123. Dumax-Vorzet, A., Roboti, P. & High, S. OST4 is a subunit of the mammalian oligosaccharyltransferase required for efficient N-glycosylation. *Journal of Cell Science* 126, 2595–2606 (2013).

124. Lee, J. *et al.* EWSR1, a multifunctional protein, regulates cellular function and aging via genetic and epigenetic pathways. *Biochimica Et Biophysica Acta. Molecular Basis of Disease* 1865, 1938–1945 (2019).

125. Li, N. *et al.* Paip1 Indicated Poor Prognosis in Cervical Cancer and Promoted Cervical Carcinogenesis. *Cancer Research and Treatment* 51, 1653–1665 (2019).

126. Kell, M. J., Ang, S. F., Pigati, L., Halpern, A. & Fölsch, H. Novel function for AP-1B during cell migration. *Molecular Biology of the Cell* 31, 2475–2493 (2020).

127. Burns, A. C. *et al.* Genome-wide gene by environment study of time spent in daylight and chronotype identifies emerging genetic architecture underlying light sensitivity. *Sleep* 46, zsac287 (2023).

128. Astuti, G. D. N. *et al.* Mutations in AGBL5, Encoding α-Tubulin Deglutamylase, Are Associated With Autosomal Recessive Retinitis Pigmentosa. *Investigative Ophthalmology & Visual Science* 57, 6180–6187 (2016).

129. Alshammari, E., Zhang, Y.-X. & Yang, Z. Mechanistic and functional extrapolation of SET and MYND domain-containing protein 2 to pancreatic cancer. *World Journal of Gastroenterology* 28, 3753–3766 (2022).

130. Bodoni, A. F. *et al.* Nicotinamide Nucleotide Transhydrogenase Is Essential for Adrenal Steroidogenesis: Clinical and In Vitro Lessons. *The Journal of Clinical Endocrinology and Metabolism* 108, 1464–1474 (2023).

131. Lemberg, M. K. Sampling the membrane: function of rhomboid-family proteins. *Trends in Cell Biology* 23, 210–217 (2013).

132. O'Mara, T. A. *et al.* Comprehensive genetic assessment of the ESR1 locus identifies a risk region for endometrial cancer. *Endocrine-related cancer* 22, 851–861 (2015).

133. Buchegger, K. *et al.* Reprimo, a Potential p53-Dependent Tumor Suppressor Gene, Is Frequently Hypermethylated in Estrogen Receptor α-Positive Breast Cancer. *International Journal of Molecular Sciences* 18, 1525 (2017).

134. Scotti, M. & Kmita, M. Recruitment of 5' Hoxa genes in the allantois is essential for proper extra-embryonic function in placental mammals. *Development (Cambridge, England)* 139, 731–739 (2012).

135. Hodgkinson, K. *et al.* GREB1 is an estrogen receptor-regulated tumour promoter that is frequently expressed in ovarian cancer. *Oncogene* 37, 5873–5886 (2018).

136. Gendron, R. L. *et al.* Abnormal uterine stromal and glandular function associated with maternal reproductive defects in Hoxa-11 null mice. *Biology of Reproduction* 56, 1097–1105 (1997).

137. Jeanne, M. *et al.* Missense variants in DPYSL5 cause a neurodevelopmental disorder with corpus callosum agenesis and cerebellar abnormalities. *American Journal of Human Genetics* 108, 951–961 (2021).

138. Deng, W. *et al.* The novel secretory protein CGREF1 inhibits the activation of AP-1 transcriptional activity and cell proliferation. *The International Journal of Biochemistry & Cell Biology* 65, 32–39 (2015).

139. Zucman-Rossi, J., Legoix, P. & Thomas, G. Identification of new members of the Gas2 and Ras families in the 22q12 chromosome region. *Genomics* 38, 247–254 (1996).

140. Harbin, L. M., Lin, N., Ueland, F. R. & Kolesar, J. M. SYNE1 Mutation Is Associated with Increased Tumor Mutation Burden and Immune Cell Infiltration in Ovarian Cancer. *International Journal of Molecular Sciences* 24, 14212 (2023).

141. Lukkahatai, N., Walitt, B., Deandrés-Galiana, E. J., Fernández-Martínez, J. L. & Saligan, L. N. A predictive algorithm to identify genes that discriminate individuals with fibromyalgia syndrome diagnosis from healthy controls. *Journal of Pain Research* 11, 2981–2990 (2018).

142. Markkula, R. *et al.* Clustering of symptoms associated with fibromyalgia in a Finnish Twin Cohort. *European Journal of Pain* 13, 744–750 (2009).

143. Dutta, D. *et al.* Heritability of the fibromyalgia phenotype varies by age. *Arthritis & rheumatology (Hoboken, N.J.)* 72, 815–823 (2020).

144. Sapkota, Y. *et al.* Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. *Nature Communications* 8, 15539 (2017).

145. Ferrari, L. F., Khomula, E. V., Araldi, D. & Levine, J. D. Marked Sexual Dimorphism in the Role of the Ryanodine Receptor in a Model of Pain Chronification in the Rat. *Scientific Reports* 6, 1–12 (2016).

146. Vacca, V. *et al.* 17beta-estradiol counteracts neuropathic pain: a behavioural, immunohistochemical and proteomic investigation on sex-related differences in mice. *Scientific Reports* 6, 1–18 (2016).

147. Chen, Q., Zhang, W., Sadana, N. & Chen, X. Estrogen receptors in pain modulation: cellular signaling. *Biology of Sex Differences* 12, 22 (2021).

148. Tu, C.-H., Lin, C.-L., Yang, S.-T., Shen, W.-C. & Chen, Y.-H. Hormonal Contraceptive Treatment May Reduce the Risk of Fibromyalgia in Women with Dysmenorrhea: A Cohort Study. *Journal of Personalized Medicine* 10, 280 (2020).

149. Stening, K. D. *et al.* Hormonal replacement therapy does not affect self-estimated pain or experimental pain responses in post-menopausal women suffering from fibromyalgia: a double-blind, randomized, placebo-controlled trial. *Rheumatology (Oxford, England)* 50, 544–551 (2011).

150. Schertzinger, M., Wesson-Sides, K., Parkitny, L. & Younger, J. Daily Fluctuations of Progesterone and Testosterone Are Associated With Fibromyalgia Pain Severity. *The Journal of Pain* 19, 410–417 (2018).

151. Dunnett, A. J., Roy, D., Stewart, A. & McPartland, J. M. The diagnosis of fibromyalgia in women may be influenced by menstrual cycle phase. *Journal of Bodywork and Movement Therapies* 11, 99–105 (2007).

152. Mostafavi, H., Spence, J. P., Naqvi, S. & Pritchard, J. K. Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nature Genetics* 55, 1866–1875 (2023).

153. Pagé, M. G., Fortier, M., Ware, M. A. & Choinière, M. As if one pain problem was not enough: prevalence and patterns of coexisting chronic pain conditions and their impact on treatment outcomes. *J Pain Res* 11, 237–254 (2018).

154. Elliott, L. T. *et al.* Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* 562, 210–216 (2018).