Immunogenetic studies in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS)

Thesis for the degree of philosophiae Doctor (PhD)

By

Riad Hajdarevic

2022

Department of Medical Genetics, Oslo University Hospital

Faculty of Medicine, University of Oslo

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo

ISBN 978-82-348-0044-3

Cover: Hanne Baadsgaard Utigard.
Print production: Graphics Center, University of Oslo.

Contents

Acknowledgement

This would have not been possible to do without the help and guidance of many people.

First and foremost I want to thank my supervisors Marte K Viken, Benedice A Lie and Ola Didrik. I have learned so much from each one of you during the last 4 years. I will continue my scientific journey with all the three of you being a part of me. Thank you!

Special thanks go to my friend and mentor Srdjan who gave me the chance to become a part of AMG and who pushed the first domino which has led to this moment.

Likewise, I want to thank Tim, Ida and Alexey who helped me with my work and took the time to give their input and support.

To all my friends from the Immgen group: Fatima, Siri, Marte, Maria, Asgeir and Anne, thank you for listening to my long rants and unasked opinions on everything.

To everyone at AMG I also say thank you. Thank you for being an amazing place to grow up as a scientist.

Thank you Lars for being there for me and having the patients with me when my hands were shaking and I had no clue what I was doing. You will always be remembered as a great friend and mentor. May you rest in peace

Hvala ti Hana sto prozivljavala sve uspone i padove ovog puta zajedno sa mnom. Ti si sve ovo omogucila i zrtvovala se vise nego ja. Koliko smo samo vikenda proveli odvojeno (bez Cileta) zbog ovog rada. Hvala ti i sto si rodila Cileta i sto si mi najbolji drug.

Hvala i tebi Zizi sto sam imao skim pricati o stvarima koje samo tebe i men zanimaju i sto si (ipak) dobar brat. Hvala Zulfeta na svemu, ovo ja tvoja pobjeda koliko i moja. Hvala Majka sto si uvjek bila uz mene koliko god blesav bio. Hvala Tetka sto si mi omogucila da dodjem ovdje i za sve ostalo. Hvala i tebi Dino (i Ana i Adrian) sto si nakon svih godina uvijek uz mene i sto iskreno gledas moje uspjehe kao svoje, uvijek mi je to znacilo.

In loving memory of my father (1963-2005)

Abbreviations

AIDs	Autoimmune disorders
APC	Anitgen presenting cell
BCR	B cell receptor
CBT	Cognitive behavior therapy
CCC	Canadian consensus criteria
CD	Crohn's disease
CeD	Celiac disease
DNA	Deoxyribonucleic acid
EBV	Epstein Bar virus
eQTL	Expression quantitative trait loci
GET	Graded exercise therapy
GWAs	Genome wise association study
HLA	Human leukocyte antigen
HWE	Hardy-Weinberg equilibrium
IBD	Inflammatory bowel diseases
IgM	Immunoglobulin M
KIR	Killer immunoglobulin like receptor
LD	Linkage disequilibrium
M3/4AChR	Muscarinic acetylcholine receptor M3/4
ME/CFS	Myalgic encephalomielytis
MG	Myasthenia gravis
MHC	Major Histocompatibility Complex
MS	Multiple sclerosis
NGS	Next generation sequencing
NK cells	Natural Killer Cells
OR	Odds ratio
P	P value
PEM	Post exertional malaise
POTS	Postural orthostatic tachycardia syndrome
PS	Psoriasis
RA	Rheumatoid arthritis
SLE	Systemic lupus erythematosus
SNP	Single nucleotide polymorphism
T1D	Type 1 diabetes
TCR	T cell receptor
TRA	T cell receptor alpha locus
UC	Ulcerative colitis
β2 AdR	Beta-2 adrenergic receptor

Gene Names

BACH2	BTB Domain And CNC Homolog 2
CACNA11	calcium voltage-gated channel subunit alpha1 C
CCR3	C-C Motif Chemokine Receptor 3
CD226	Cluster of Differentiation 226
CEP72	Centrosomal Protein 72
CTLA4	cytotoxic T-lymphocyte-associated protein 4
EPHA7	EPH Receptor A7
FUT2	Fucosyltransferase 2
IFIH1	Interferon Induced With Helicase C Domain 1
IGFBP1	Insulin Like Growth Factor Binding Protein 1
IGFBP3	Insulin Like Growth Factor Binding Protein 3
IL18RAP	Interleukin 18 Receptor Accessory Protein
IL18RAP	Interleukin 18 Receptor Accessory Protein
IL21	Interleukin 21
IL23R	Interleukin 23 Receptor
IL2RA	Interleukin 2 Receptor Subunit Alpha
IZUMO1	Izumo sperm-egg fusion 1
LINC00333	LINC00333
LPP	LIM Domain Containing Preferred Translocation Partner In Lipoma
MAMSTR	MEF2 Activating Motif And SAP Domain Containing Transcriptional Regulator
MMEL1	Membrane metallo-endopeptidase-like 1
NOX3	NADPH Oxidase 3
NR3C1	Nuclear Receptor Subfamily 3 Group C Member 1
OLIG3	Oligodendrocyte Transcription Factor 3
PRKCQ	Protein Kinase C Theta
PTPN2	Protein Tyrosine Phosphatase Non-Receptor Type 2
PTPN22	protein tyrosine phosphatases
RGS1	Regulator Of G Protein Signaling 1
RIN3	Ras And Rab Interactor 3
SH2B3	SH2B Adaptor Protein 3
SHANK3	SH3 And Multiple Ankyrin Repeat Domains 3
SKAP1	Src Kinase Associated Phosphoprotein 1
STAB2	Stabilin 2
STAT4	Signal transducer and activator of transcription 4
STMN3	Stathmin 3
TAGAP	T Cell Activation RhoGTPase Activating Protein
TNFAIP3	TNF Alpha Induced Protein 3
TNFSF14	TNF Superfamily Member 14
TPPP	Tubulin Polymerization Promoting Protein
TRA	T cell receptor alpha locus
UBASH3A	Ubiquitin Associated And SH3 Domain Containing A
ZBTB46	Zinc Finger And BTB Domain Containing 46

List of publications

Paper I

Riad Hajdarevic, Asgeir Lande, Ingrid Rekeland, Anne Rydland, Elin B. Strand, Daisy D. Sosa, Lisa E Creary, Olav Mella, Torstein Egeland, Ola D. Saugstad, Øystein Fluge, Benedicte A. Lie, Marte K. Viken: *Fine mapping of the major histocompatibility complex (MHC) in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) suggests involvement of both HLA class I and class II loci*, Brain, Behavior, Immunity,2021, https://doi.org/10.1016/j.bbi.2021.08.219

Paper II

Riad Hajdarevic, Asgeir Lande, Jesper Mehlsen, Anne Rydland, Daisy D. Sosa, Elin B. Strand ,Olav Mella, Flemming Pociot, Øystein Fluge, Benedicte A. Lie1, Marte K. Viken **:** *Genetic association study in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) identifies several potential risk loci (Submitted)*

Paper III

Marthe Ueland, **Riad Hajdarevic**, Olav Mella, Elin B. Strand, Daisy D. Sosa, Ola D. Saugstad, Øystein Fluge, Benedicte A. Lie1, Marte K. Viken*: No replication of previously reported association with genetic variants in the T cell receptor alpha (TRA) locus for myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) (Submitted)*

Summary in Norwegian

Myalgisk encefalopati/kronisk utmattelsessyndrom (ME/CFS) er en kronisk og invalidiserende tilstand som affiserer ca 0,1-0,2% av den generelle befolkningen. Hoved symptomene er vedvarende utmattelse som gir betydelig reduksjon av funksjonsnivået i forhold til tidligere, langvarig forverring etter anstrengelse (på engelsk kalt PEM, «postexertional malaise») og kognitiv dysfunksjon hos pasienter med ME/CFS. De fleste symptomene på ME/CFS er ikke sykdomsspesifikke. Samtidig mangler det både biomarkører og diagnostiske tester for sykdommen, noe som vanskeliggjør diagnostiseringen av tilstanden. Mer enn 20 forskjellige pasientklassifiseringer og diagnostiske kriterier har dukket opp i løpet av de siste fire tiårene. Pasientpopulasjonen kan dermed være ganske heterogen med tanke på kliniske symptomer og i hvor stor grad sykdommen påvirker livskvalitet

Det eksisterer flere ulike teorier som tar sikte på å forklare årsakene til sykdomsutvikling. I denne oppgaven har vi tatt utgangspunkt i de økende bevisene for en immunologisk bakgrunn for ME/CFS sykdomsutvikling. Flere studier har pekt mot endrede NK-celler, autoantistoffer og T-celleavvik hos ME/CFS-pasienter. I tillegg rapporterte flere genetiske studier signifikante assosiasjoner i forskjellige immunologisk relevante gener. De fleste av disse tidligere studiene på feltet har vært suboptimale og inkluderte enten få pasienter, ulike diagnostiske kriterier eller begge deler. Derfor hadde vi som mål å oppnå en bedre forståelse av rollen immunologiske gener kan ha i henhold til sykdommen gitt tidligere observasjoner i immunologiske og genetiske studier. For å gjøre dette brukte vi kjente strategier fra genetiske studier på autoimmun sykdom for å studere ME/CFS. Vi brukte streng kvalitetskontroll og inkluderte, så vidt vi vet, den største kohorten diagnostisert med de kanadiske konsensuskriteriene.

I **artikkel I** var hovedmålet å følge opp tidligere utført arbeid av gruppen vår som rapporterte assosiasjoner mellom ME/CFS og HLA-C:07:04 og HLA-DQB1:03:03 alleler. HLA (human leukocytt antigen) regionen består av en rekke immunologisk relevante gener i tillegg til HLA genene, og det er omfattende og kompleks koblingsulikevekt i regionen. De tidligere observerte assosiasjonssignalene i HLA-regionen ble forsøkt finkartlagt ved å genotype fem ekstra klassiske HLA-loci og 5 342 SNPer (enkelt nukleotid varianter) hos 427 norske ME/CFS-pasienter, diagnostisert i henhold til de kanadiske konsensuskriteriene, og 480 friske norske kontroller. Analysene avdekket to uavhengige assosiasjonssignaler ($p \leq 0{,}001$) representert med de genetiske variantene rs4711249 i HLA klasse I-regionen og rs9275582 i HLA-klasse II-regionen. Det primære assosiasjonssignalet i HLA klasse II-regionen var lokalisert i HLA-DQ genregionen, mest sannsynlig på grunn av HLA-DQB1 genet. Spesielt aminosyreposisjon 57 (asparaginsyre/alanin) i peptidbindingsgropen til HLA-DQB1, eller en SNP oppstrøms for HLA-DQB1 så ut til å forklare assosiasjonssignalet vi observerte i HLA-klasse II regionen. I HLA klasse I-regionen var det antatte primære lokuset ikke like tydelig og kan muligens ligge utenfor de klassiske HLA-genene (assosiasjonssignalet spenner over flere gener DDR1, GTF2H4, VARS2, SFTA2 og DPCR1) med ekspresjonsnivåer påvirket av ME/CFS assosiert SNP genotype. Interessant nok observerte vi også at nesten 62 % av pasientene som responderte på cyklofosfamidbehandling for ME/CFS hadde enten rs4711249-risikoallelet og/eller DQB1*03:03 mot 12 % av pasientene som ikke responderte på behandlingen. Våre funn tyder på involvering av HLA-regionen, og spesielt HLA-DQB1-genet, i ME/CFS. Selv om vår studie er den største per i dag er det fortsatt en relativt liten

studie i genetisk sammenheng, våre funnene må derfor replikeres i mye større kohorter. Spesielt er det nødvendig å undersøke involveringen av HLA-DQB1, et gen som inneholder alleler som gir økt risiko for flere etablerte autoimmune sykdommer slik som cøliaki.

I **artikkel II** hadde vi som mål å undersøke immunologisk relevante gener ved å bruke en genotypings array (iChip) rettet mot immunologisk gen-regioner tidligere assosiert med diverse autoimmune sykdommer. I tillegg til den norsk kohort på 427 ME/CFS-pasienter (de kanadiske konsensuskriteriene) analysert vi også data fra to replikasjons kohorter, en danskt på 460 ME/CFS-pasienter (de kanadiske konsensuskriteriene) og et datasett fra den britiske biobanken på 2105 selvrapporterte CFS-pasienter. Så vidt vi vet, er dette den første ME/CFS assosiasjonsstudie av denne størrelsen som inkluderer mer enn 2900 pasienter (hvor av 887 er diagnostisert i henhold til de kanadiske konsensuskriteriene). Vi fant ingen ME/CFS-risikovarianter med et såkalt helgenom statistisk signifikansnivå ($p<5x10^{-8}$), men vi identifiserte seks genregioner (TPPP, LINC00333, RIN3. IGFBP1/IGFBP3, IZUMO1/MAMSTR og ZBTB46/STMN3) som viste tegn til mulig assosiasjon med ME/CFS og som krever videreoppfølgingi fremtidige genetiske studier for å kunne vurdere om de er reelle funn eller ikke. Interessant nok uttrykkes disse genene i sykdomsrelevant vev, f.eks. hjerne, nerve, skjelettmuskulatur og blod, inkludert immunceller som undergrupper av T-celler, B-celler, NK-celler og monocytter. Videre er flere av de ME/CFS-assosierte SNP-genotypene assosiert med differensielle ekspresjonsnivåer av disse genene. Selv om vi ikke kunne identifisere statistisk overbevisende assosiasjoner med genetiske varianter på tvers av de tre kohortene, mener vi at våre datasett og analysere representerer et viktig steg i forskningen på ME/CFS. Vår studie viser at for den fremtidig forståelse av genetiske varianter som kan gi økt sårbarhet for å utvikle kreves det enda større studier.

I **artikkel III** ønsket vi å undersøke tidligere funn fra en helgenom assosiasjonsstudie på 42 ME/CFS-pasienter som rapporterte signifikant assosiasjon med to SNPer i T-cellereseptor alfa (TRA) locus (P-verdi <5x10-8). For å forsøke å replikere disse tidligere funnene brukte vi en stor norsk ME/CFS-kohort (409 tilfeller og 810 kontroller) og data fra den britiske biobanken (2105 tilfeller og 4786 kontroller). Vi undersøkte en rekke SNP-er i TRA-lokuset, inkludert de to tidligere ME/CFS-assosierte variantene, rs11157573 og rs17255510. Ingen statistisk signifikante assosiasjoner ble observert i den norske kohorten, heller ikke for de to tidligere rapporterte SNPene i noen av kohortene. Riktignok viste andre SNPer svake tegn på assosiasjon (P-verdi <0,05) i den britiske biobankkohorten og metaanalyser av norske og britiske biobankkohorter, men ingen overlevde korreksjon for multiple tester. Vi kunne altså ikke bekrefte assosiasjoner med genetiske varianter i TRA-lokuset i denne studien.

Summary in English

Myalgic encephalopathy/chronic fatigue syndrome (ME/CFS) is a chronic and debilitating disease that affects about 0.1-0.2% of the general population. The core symptoms are persistent debilitating fatigue, post-exertional malaise (PEM) and cognitive dysfunction. Most symptoms of ME/CFS are not disease specific. Additionally, there is a lack of both biomarkers and diagnostic tests for the disease, which makes accurate diagnosis difficult. More than 20 different patient classifications and diagnostic criteria have emerged over the last four decades. Due to this, the patient population can be quite heterogeneous in terms of clinical symptoms and the extent to which the disease impacts quality of life.

There are several different theories that aim to explain the disease development of ME/CFS. In this thesis, we have taken as our starting point the growing evidence for an immunological background for ME/CFS pathogenesis. Several studies have pointed to altered NK cells, autoantibodies and T cell abnormalities in ME/CFS patients. In addition, several genetic studies reported significant associations in various immunologically relevant genes. Most of these previous studies have been suboptimal and included heterogeneous patient populations and/or few patients in total. Therefore, we aimed to gain a better understanding of the role of immunologically relevant genes and disease development of ME/CFS. To do this, we employed known strategies from genetic studies in autoimmune disease and applied them to ME/CFS. We used strict quality control and included, to the best of our knowledge, the largest cohort diagnosed with the Canadian consensus criteria.

In paper I, the main goal was to follow up previously performed work by our group that reported associations between ME/CFS and HLA-C: 07: 04 and HLA-DQB1: 03: 03 alleles. The HLA (human leukocyte antigen) region consists a multitude of immunologically relevant genes in addition to the HLA genes, and there is extensive and complex linkage disequilibrium (LD) in the region. The previously observed association signals in the HLA region were fine-mapped by genotyping five additional classical HLA loci and 5,342 SNPs (single nucleotide variants) in 427 Norwegian ME/CFS patients, diagnosed according to the Canadian consensus criteria, and 480 healthy Norwegian controls. The analysis revealed two independent association signals ($p \leq 0.001$) represented by the genetic variants rs4711249 in the HLA class I region and rs9275582 in the HLA class II region. The primary association signal in the HLA class II region was located in the vicinity of the HLA-DQ genetic region, most likely due to the HLA-DQB1 gene. In particular, amino acid position 57 (aspartic acid / alanine) in the peptide binding pit of HLA-DQB1, or an SNP upstream of HLA-DQB1 seemed to explain the association signal we observed in the HLA class II region. In the HLA class I region, the putative primary locus was not as clear and could possibly lie outside the classical HLA genes (the association signal spans several genes *DDR1*, *GTF2H4*, *VARS2*, *SFTA2* and *DPCR1*) with expression levels influenced by the ME/CFS associated SNP genotypes. Interestingly, we also observed that > 60% of the patients who responded to cyclophosphamide treatment for ME/CFS had either the rs4711249 risk allele and/or DQB1* 03:03 versus 12% of the patients who did not respond to the treatment. Our findings suggest the involvement of the HLA region, and in particular the HLA-DQB1 gene, in ME/CFS. Although our study is the largest to date, it is still a relatively small study in the context of genetic studies. Our findings need to be replicated in much larger, statistically more representative, cohorts. In particular, it is necessary to investigate the involvement of HLA-

DQB1, a gene that contains alleles that increase the risk of several established autoimmune diseases such as celiac disease.

In paper II, we aimed to investigate immunologically relevant genes using a genotyping array (iChip) targeting immunological gene regions previously associated with different autoimmune diseases. In addition to the Norwegian cohort of 427 ME/CFS patients (the Canadian consensus criteria), we also analyzed data from two replication cohorts, a Danish one of 460 ME/CFS patients (Canadian consensus criteria) and a data set from the UK Biobank of 2105 self-reported CFS patients. To the best of our knowledge, this is the first ME/CFS genetic association study of this magnitude and it included more than 2,900 patients in total (of whom 887 are diagnosed according to Canadian consensus criteria). We found no ME/CFS risk variants with a genome wide significance level ($p < 5x10-8$), but we identified six gene regions (*TPPP*, *LINC00333*, *RIN3*. *IGFBP*/*IGFBP3*, *IZUMO1*/*MAMSTR* and *ZBTB46*/*STMN3*) with possible association with ME/CFS which require further follow-up in future studies in order to assess whether they are real findings or not. Interestingly, these genes are expressed in disease-relevant tissue, e.g. brain, nerve, skeletal muscle and blood, including immune cells (subgroups of T cells, B cells, NK cells and monocytes). Furthermore, several of the ME/CFS associated SNP genotypes are associated with differential expression levels of these genes. Although we could not identify statistically convincing associations with genetic variants across the three cohorts, we believe that our data sets and analysis represent an important step in the ME/CFS research field. Our study demonstrated that for the future understanding of the genetic architecture of ME/CFS much larger studies are required to established reliable associations.

In paper III, we wanted to investigate previous findings from a genome wide association study of 42 ME/CFS patients who reported significant association with two SNPs in the T cell receptor alpha (TRA) locus (P-value $<5x10-8$). In order to replicate these previously reported findings, we used a large Norwegian ME/CFS cohort (409 cases and 810 controls) and data from the UK Biobank (2105 cases and 4786 controls). We examined a number of SNPs in the TRA locus, including the two previous ME/CFS-associated variants, rs11157573 and rs17255510. No statistically significant associations were observed in either the Norwegian cohort or UK biobank cohorts. Nevertheless, other SNPs in the region showed weak signs of association (P-value <0.05) in the UK Biobank cohort and meta-analyzes of Norwegian and UK Biobank cohorts, but did not remain associated after applying correction for multiple testing. Thus, we could not confirm associations with genetic variants in the TRA locus in this study.

1 Introduction

1.1 Myalgic encephalomyelitis /chronic fatigue syndrome (ME/CFS)

Myalgic encephalomyelitis /chronic fatigue syndrome (ME/CFS) is a debilitating and chronic disease of unknown pathogenesis that affects 0.1%–0.2% (using the Canadian consensus criteria) of the general population [1]. The hallmarks of the condition are persistent fatigue, post-exertional malaise (PEM) and cognitive dysfunction over a period of at least 6 months [1]. Currently, neither adequate treatments nor biomarkers are available. The lack of biomarkers is not only a clinical disadvantage but also hinders conducting standardized research in the field, which could, in consequence, lead to the development of effective treatment options.

Historically, different terms have been coined in order to name unexplained fatigue. In this context, each of the names reflects diagnostic or other biases. Terms like Neurasthenia, epidemic neuromyasthenia, Royal Free disease, Akureyri disease, chronic Epstein-Bar virus (EBV) syndrome, and fibromyalgia have all been used to describe states which symptomatically overlap with ME/CFS [2, 3]

The debate on which symptoms to include in diagnosing ME/CFS is ongoing and constantly under revision. There are more than 25 case definitions for ME/CFS thus far [4]. All the proposed diagnostic criteria could, in practice, be categorized according to what primary symptom they focus on [4]. In no particular order, the eight most frequently used in research are: Holmes, Ramsay, Oxford, Fukuda, Australian, Canadian consensus criteria (CCC) which are used in this work, International consensus criteria (ICC), and systemic exertion intolerance disease (SEID) [4]. The different diagnostic criteria might be one reason why the prevalence estimates vary from roughly 0.2 % to 2% [4, 5].

The ME/CFS research community has proposed many theories on the aetiology and disease development of ME/CFS. This diversity has sparked huge scientific debate, which has generated competing views on the topic. One side of the community puts ME/CFS in a strict psychosocial model whilst the other part adheres to a more biomedical underlying cause of ME/CFS. To be fair, this division is slowly but surely dissolving under the pressure of new evidence suggesting a very complex and heterogeneous disease picture encompassing, at least in part, both sides of the discussion.

1.1.1 Diagnostic criteria for ME/CFS

Even after years of ME/CFS research, it has been shown to be difficult to properly assess ME/CFS in a clinical or laboratory setting. Many sets of criteria have been developed with the intention to properly assess ME/CFS thereby minimizing possible misdiagnosis with other illnesses [6]. However, the objective of a particular diagnostic criteria is, likewise, very important. On one hand, physicians and health care providers would, most probably, want to make sure all potential cases are included. On the other hand, this approach is less beneficial when trying to conduct proper research since the primary objective is a homogenous patient population. It is, therefore, important to keep this in mind when evaluating any set of diagnostic criteria.

Without proper biomarkers, ME/CFS diagnosis has been subjugated to be diagnosed mostly by an exclusion approach. The clinical picture of ME/CFS can easily be confused with many other illnesses due to the commonness of most symptoms of ME/CFS in other pathologies [4, 7].

Research has identified many pathological patterns in ME/CFS but none of those are enough to construct a unified diagnostic test that would definite point towards a ME/CFS diagnosis [8].

The term Chronic fatigue syndrome was used for the first time in the late 1980s by Holmes et al. [9]. Holmes hence proposed the first diagnostic criteria. These so called Holmes criteria were presented in the same year (1988) [10]. To date, 25 case definitions like the Fukuda et al. classification [11], the CCC [1]and the Oxford criteria [12] have emerged[4]. Each diagnostic criteria is made up of categories such as: required conditions, inclusions and exclusionary symptoms/disorders.

The recent review of Lim et al. has identified the Fukuda criteria to be the most commonly cited in ME/CFS research [4]

Looking into the specifics of the Fukuda criteria the following parameters apply to assess ME/CFS:

1. Clinically assessed, unexplained, chronic or relapsing fatigue that is of new or with known onset. The fatigue is not the result of ongoing exertion or is not substantially improved by rest and results in a major reduction in previous levels of everyday activities [11]

2. The parallel occurrence of four or more core symptoms, all of which must have been experienced on a persistent or reoccurring basis during six or more months of illness and must not have predated the fatigue [11]

The main symptoms are defined as:

- Persistent fatigue of 6 with months duration
- Fatigue affects daily activities
- Fatigue is not alleviated by rest
- Fatigue is new or definite in onset
- Fatigue is not explained by any other condition
- Unrefreshing sleep [11]

Additional symptoms (4 need to be present)

- Impaired memory
- Sore throat
- Tender lymph nodes
- Muscle pain
- Joint paint
- New headache
- PEM
- Unrefreshing sleep [11]

The CCC, used in this work, have several different diagnostic requirements than the Fukuda criteria and state the following:

1. Symptom severity must result in a 50% or greater reduction of a patient's premorbid activity level for a diagnosis. [1]

2. At least one symptom from the four major groupings are required:

- **Post-Exertional Neuroimmune Exhaustion(PEM)**
- **Neurological impairment grouping:**
- (1) neurocognitive impairments, (2) pain, (3) sleep disturbance, and (4) neurosensory, perceptual and motor disturbance
- **Immune, Gastro-intestinal and Genitourinary Impairments grouping:**

- (1) flu-like symptoms, (2) susceptibility to viral infections with prolonged recovery periods (3) gastro-intestinal tract symptoms (4) genitourinary symptoms and (5) sensitivities to food, medications, odours, or chemicals
- **Energy Production/transportation impairments grouping:**
- (1) cardiovascular, (2) respiratory, (3) loss of thermostatic stability, and (4) intolerance of extremes of temperature. [1]

When comparing the Fukuda and the CCC, it is evident that the Fukuda criteria do not insist on core symptoms of ME/CFS, like post-exertional malaise (PEM) and neurocognitive symptoms. The Fukuda criteria have been criticized for being too inclusive, and thereby being in danger of including patients whose symptoms could be of a psychiatric origin[13]. Since some of the symptoms intersect with the symptoms of depression, patients with depression may be misdiagnosed as ME/CFS patients by the Fukuda definition [14]. In contrast, the CCC requires for a diagnosis of ME/CFS the presence of core symptoms like fatigue, sleep dysfunction, PEM as well as pain and several minor symptoms [1, 4].

In addition to the CCC [1] and the Fukuda criteria [11] the Oxford criteria [12] are widely used in certain areas of inquiry. They are regarded as more inclusive since they require only mild fatigue severity in contrast to the moderate and severe fatigue required in the Fukuda and CCC [15]. The Oxford criteria define Chronic Fatigue Syndrome (CFS) if mild to severe symptoms of fatigue, sleep disturbance and myalgia are present [12]. However, the presence or absence of other complaints is not required. Likewise, states and comorbidities that are otherwise excluded in most diagnostic criteria (e.g. psychiatric diseases) are permitted by the Oxford criteria [12].

The main factors that have modelled ME/CFS case definitions are aetiology, pathophysiology, and exclusionary disorders. This affected the specification of case definitions(Table 1) [4].

The above mentioned diagnostic criteria (Fukuda, CCC and Oxford) all have their application in certain clinical settings [4, 15, 16].

Table 1. Overview of symptoms needed for ME/CFS diagnosis in accordance to difference diagnostic criteria

Diagnostic criteria	**CDC/Fukuda criteria** [11]	**Canadian Consensus Criteria (2003) (ME/CFS)** [1]	**Oxford criteria(CFS)** [12]	**NICE guidelines (2021) [17]**
Symptoms	•Persistent fatigue of 6 months duration •Fatigue affects daily activities • Fatigue is not alleviated by rest • Fatigue is new or definite in onset •Fatigue is not explained by any other condition •Unrefreshing sleep **Additional symptoms (4 need to be present)** •Impaired memory •Sore throat •Tender lymph nodes •Muscle pain •Joint paint •New headache •PEM •Unrefreshing sleep	One symptom from each of the four groupings are required: •**Post-Exertional Neuroimmune Exhaustion (PEM)** •**Neurological impairment grouping:** (1) neurocognitive impairments, (2) pain, (3) sleep disturbance, and (4) neurosensory, perceptual and motor disturbance •**Immune, Gastro-intestinal and Genitourinary Impairments grouping:** (1) flu-like symptoms, (2) susceptibility to viral infections with prolonged recovery periods (3) gastro-intestinal tract symptoms (4) genitourinary symptoms and (5) sensitivities to food, medications, odors, or chemicals •**Energy Production/transportation impairments grouping:**(1) cardiovascular, (2) respiratory ,(3) loss of thermostatic stability, and (4) intolerance of temperature.	Fatigue, definite onset, severe and disabling, affects physical and mental functioning (duration > 6 months, symptoms present > 50% of the time)	Debilitating fatigue that is worsened by activity Post-exertional malaise after activity Unrefreshing sleep or sleep disturbance (or both) Cognitive difficulties (sometimes described as 'brain fog')

Nevertheless, the lack of a universally accepted set of criteria has made it very hard to compare research results. Like with any complex phenotype it is of utmost importance to make the study population as homogenous as possible to yield reproducible results. That is why some researchers consider the CCC (used in this work) to be more superior since they exclude a substantial amount of patients which would be diagnosed with ME/CFS otherwise [18]. However, as mentioned before the Fukuda are still the most cited diagnostic criteria in ME/CFS research [4].

1.1.2 Proposed pathogenesis of ME/CFS

The pathogenesis of ME/CFS is a highly debated and sometimes polarizing topic**.** Since its definition, ME/CFS has been trapped between biomedicine and psychiatry [19]

One suggested disease model for ME/CFS is the biopsychosocial model [20]. Sometime in the 1980's, physicians fused neurasthenia, hysteria, and somatoform illness, to reconstitute ME as chronic fatigue syndrome (CFS)[9]. Some scientists argue that CFS is best understood using a biopsychosocial model e.g. being triggered by viral triggers, but maintained by the sufferers' mind and social environment. [20]. Fatigue and impairment are considered to be the end result of psycho-social factors predominantly. According to this model, attributing complaints to a biological cause negatively influences activity status, which in turn has a negative impact on the severity of fatigue and impairment of the patients [20]. Therefore, interventions like cognitive behavioural therapy (CBT) and graded exercise therapy (GET) are proposed in order to change the patient's notion of fatigue and improve wellbeing. It is advised to not view their perceived fatigue as sensations of some physical problem, but rather as unhealthy or obsessive thought processes [21]. The evidence for this model can be described as modest, at least when looking at it from a more stringent, biological point of view [22, 23]. The recently revised guidelines for ME/CFS treatment and care from the National Institute for Health and Care Excellence in the UK (NICE) now discourage the use of GET as a treatment option, considering it as scientifically unjustified, and CBT is only recommended as supportive therapy. [17] (Table 1).

A different approach towards understanding ME/CFS has been gaining ground in recent years. Even though this disease model has no official name like the biopsychosocial one mentioned before, it can be put under the umbrella term of the biological model of ME/CFS. A growing body of research suggests that an intricate interplay of the nervous, endocrine, metabolic and immune systems might play a crucial role in ME/CFS development [24, 25, 26, 27, 28].

1.1.3 Immunological studies in ME/CFS

The dysregulation or dysfunction of the immune system is proposed to be involved in ME/CFS disease development [24]. Hence, a multitude of studies has been conducted in order to investigate this hypothesis.

Acute viral infections have been a well-established trigger for autoimmune disorders (AIDs) [29]. A systematic review by Rasa et al. investigated numerous studies showing ME/CFS onset after viral episodes with various human herpesviruses, enteroviruses, the human parvovirus B19, Retroviruses and the Ross River virus [30]. They concluded that the currently available data is still unclear of whether chronic viral infection causes ME/CFS. However, they also report potential viral involvement for some subgroups of ME/CFS patients. Of the 29 reviewed studies investigating human herpes virus involvement in ME/CFS (or in some cases post-viral fatigue), 12 showed some correlation with the virus and ME/CFS while 17 failed to do that. Likewise, of the 16 studies investigating various enteroviruses, 9 showed a positive correlation with the disease while 7 did not. It is important to stress that most of the studies reviewed had several limitations like small sample size and diverging diagnostic criteria. [30]

Besides the viral triggers, ME/CFS is reported to share a substantial amount of symptoms and features with AIDs [24].

The role of natural killer (NK) cells has long been suggested to play a crucial role in autoimmune disease development [31]. NK cells could potentially be involved in AIDs given their ability of autoreactivity or through their interaction with other cell types (T cells, dendritic cells and macrophages) [32]. A systemic review by Eaton-Fitch et al. investigated 17 NK cell ME/CFS studies conducted from 1994 to 2018. The studies reviewed analyzed NK cell cytotoxicity, NK cell phenotype and receptor profiles, NK cell cytokine production, NK cell lytic protein levels and NK cell degranulation in ME/CFS [33]. The only consistently reported finding was impaired NK cell cytotoxicity while the other findings (NK cell phenotype and receptor profiles, NK cell cytokine production, NK cell lytic protein levels and NK cell degranulation) differed between studies [33]. However, it has to be mentioned that the average number of ME/CFS cases across all studies was 48. In addition, 15 out of the 17 used the Fukuda criteria for diagnosis [33].

A key feature of autoimmune disorders is the generation of autoantibodies. Several studies reported autoantibodies presence in ME/CFS [34-38]. Most notably, Loebel et al. analyzed 268 patients (CCC) and 104 controls and reported significantly elevated levels of antibodies against β adrenergic and muscarinic cholinergic receptors in a subset of patients with ME/CFS [38]

Other well-known key contributors in the pathogenesis of autoimmune diseases are cytokines [39]. Cytokines are a category of small proteins which include interferons, interleukins, chemokines, lymphokines, and tumour necrosis factors. Several studies report changed serum cytokine levels in patients with ME/CFS [40, 41, 42]. Some limitations of cytokine studies are their natural oscillation as well as their tissue specificity. Nevertheless, IFN- γ, IL12p40 and CSF1 are present in two of the three referenced studies [40, 41]. Like with previous studies, most of the cytokine studies used the Fukuda criteria.

1.1.4 Metabolic studies in ME/CFS

When looking at the clinical picture of ME/CFS (fatigue, malaise etc.) energy metabolism seems to be an obvious place to investigate disease development. Core metabolic processes like glycolysis, respiration, the tricarboxylic acid and urea cycles, glycogen catabolism, oxidative phosphorylation, and the supremacy of adenosine triphosphate in energy-transfer reactions are all well understood for several decades [43]. The primary purpose of all those processes is the conversion of macronutrients into essential cellular building blocks and adenosine triphosphate [43]. In aerobic metabolism carbohydrates, fat, and amino acids are broken down from pyruvate to ccetyl coenzyme A– in the presence of pyruvate dehydrogenase - which is fed into the tricarboxylic acid cycle. Through a redox process energy in the form of adenosine triphosphate is produced in the mitochondria with oxygen as electron acceptors. When lack of oxygen, an anaerobic state, and pyruvate is transformed to lactic acid, and adenosine triphosphate is produced without oxygen as an electron acceptor. Recent studies have reported alterations in mitochondrial mass function, impaired pyruvate dehydrogenase function and common and variable metabolic phenotypes in ME/CFS [25, 27, 44]. Most notably, Fluge et al. reported that the serum amino acid profile was changed in ME/CFS (200 CCC diagnosed cases and 102 controls) suggesting impaired mitochondrial pyruvate oxidation [27]. This is particularly interesting since cellular metabolism plays a key role in supporting immune cell maintenance and development. In primary biliary cirrhosis,

which is marked by severe fatigue, autoantibodies target pyruvate dehydrogenase complex components, causing inflammatory processes [45]

1.1.5 Prevalence, Demographics and heritability in ME/CFS

A recent meta-analysis by Lim et al. looked into 45 articles and 56 prevalence datasets, which included more than a million participants [46]. They estimate the prevalence of ME/CFS to be 0.89% according to the most commonly used diagnostic criteria in the studies they analyzed (Fukuda), with women around 2 folds higher than men regardless of diagnostic criteria. [46]. It might not be surprising that they observed diverging prevalence rates by case definitions and diagnostic methods[46].

However, besides all the variance in ME/CFS prevalence calculations, women still are the most affected by this disease across all data sets [46]. A population-based study by the Norwegian Institute of Public Health demonstrated a 3.2-fold female predominance [48]. The difference in gender proportion in ME/CFS could be related to biological factors, like sex-specific hormones or immunologic responses. Similar gender-related differences in prevalence have been observed in established autoimmune diseases [47, 49].

The heritability of ME/CFS was previously reported [50, 51]. Adult twin studies in ME/CFS have demonstrated much higher occurrence levels in monozygotic twins compared to dizygotic twins (concordance rates of 55% in monozygotic and 19% in dizygotic twins) [52, 53, 54]. External contributors like infections or upbringing during childhood is mostly similar within monozygotic and dizygotic twin pairs. Differences between monozygotic twins (100% shared DNA) and dizygotic twins (on average 50% shared DNA) can be used to estimate the contribution made by genetic heritability. The difference in concordance rates increases with increasingly stringent case definitions in adults [1]. This indicates that there is a main ME/CFS group among whom genetic factors play an important role.

1.1.6 Genetic associations in ME/CFS

A study by Smith et al. performed a genome-wide association study (GWAS) for 40 cases (defined using Fukuda criteria) and 40 control subjects. They reported 65 SNPs (single nucleotide polymorphisms) as being associated at a suggestive level of $p < 0.001$ [55]. However, none reached the GWAS significance threshold of $P < 5 \times 10^{-8}$. This inability to

detect GWAS significance is expected since GWAS with a low number of cases are not well enough powered to identify alleles with small or medium effect sizes.

Another GWAS performed by Schlauch et al. used a similarly small patient sample (42 cases, defined using CCC, and 38 controls) and reported 299 genetic variants as suggestively associated ($P < 1 \times 10^{-5}$) with ME/CFS [56]. Likewise, they reported fifteen variants as GWAS significant at $P < 5 \times 10^{-8}$. Five of those 15 are in the non-coding part of the T cell receptor alpha (TRA) locus, hence, pointing towards immune involvement in the disease. However, they also report no overlap between their findings and the 65 SNPs reported by Smith et al. [55].

Another approach on patient collection and analysis was taken by Perez et al. [57]. In this study 383 ME/CFS patients underwent DNA testing using the company 23andMe. Participants with ME/CFS were self-reported via an online pre-screening questionnaire. However, patients were not compared against a healthy control data set. Instead, they used publicly available allele frequency data via Kaviar [58] and compared it to their cohort. They report several biological relevant pathways to be involved in ME/CFS [57].

Besides GWAS, several candidate gene studies have been performed for ME/CFS. Goertzel et al. analyzed nine SNPs in the gene *NR3C1* (40 patients and 42 controls) [59]. Four SNPs (rs1866388, rs2918419, rs860458 and rs6188) were reportedly associated using a threshold of $P < 0.05$. A more recent study by Steiner et.al has also reported significant associations with two out of six tested candidate SNPs [60]. They report SNPs in the *PTPN22* and *CTLA4* genes to reach a significance of $P < 0.001$.

Several studies, including our work, investigated associations between ME/CFS and the human leukocyte antigen (HLA) genes [61, 62, 63, 64, 65, 66]. Prior to our studies [63, 64] associations with HLA alleles *HLA-DQA1*01, HLA-DRB1*13:01* and *HLA-DQB1*06:02* have been reported [62, 65, 66]. A majority of these studies incorporate <50 patients and are therefore statistically underpowered for detecting associations of moderate effect size.

1.2 Autoimmunity and Autoimmune disease

1.2.1 Autoimmunity

One of the main tasks of the immune system is the ability to induce an inflammatory response to various pathogens and toxins while, at the same time, avoiding damage to own tissues and organs. Immune tolerance prevents the immune system from reacting to self-molecules that could possibly induce an immune response in the body. Immune tolerance is often divided into central and peripheral tolerance [67]. For T cells, central tolerance is established by neutralizing autoreactive T cell clones in the thymus before they develop into fully immunocompetent cells. In the thymic cortex, positive selection occurs before the cells enter the circulation [68]. In this process, T cells in the thymus present self-peptide fragments on their own MHC/HLA class I and II molecules. Apoptosis is induced in the cell if the T cell receptor (TCR) cannot bind. However, if the TCR binds to the MHC/HLA complexes on the thymic cells, the T cell does not undergo apoptosis and is thus positively selected [68]. Accordingly, if cells react strongly to self-antigens they undergo apoptosis i.e. negative selection[68].

While central tolerance for T cells is established in the thymus, B cells undergo negative selection in the bone marrow. Maturing B cells first express B cell receptors (BCR) in the form of immunoglobulin M (IgM) and are considered immature B cells [69]. Immature B cells in the bone marrow undergo negative selection if they bind self-peptides. The self-reacting B cell can undergo receptor editing in order to change its binding specificity and avoid apoptosis [69]. Receptor editing is a process of somatic gene rearrangement (also known as V(D)J recombination) in order to give B cells with autoreactive receptors a second survival chance [70].

After exiting into the periphery, T cells and B cells are checked again by peripheral mechanisms. In the case of T cells, self-reactive cells that avoid thymic negative selection are incapacitated in the peripheral tissue by either clonal deletion, conversion to regulatory T cells (Tregs) or induction of anergy [71]. In this process, most self-reactive cells are neutralized [71, 72]. For B cells the peripheral tolerance is much less understood and is largely mediated by B cell dependence on T cell help [70, 73, 74]. Recognition of the self plays a crucial part in shaping the repertoires of immune receptors on both T and B cells [75].

Autoimmunity can have a wide range of manifestations. Some levels of autoimmunity are, as mentioned, essential for lymphocyte selection and immune functioning. Additionally, autoimmunity can manifest itself as circulating autoantibodies and minor tissue intrusion

without pathological outcomes [76, 77]. However, autoimmunity with immune-mediated organ injury (autoimmune diseases) can cause significant morbidity and even death [72, 78, 79].

In regards to ME/CFS several studies have reported elevated autoantibodies as well as T cell alterations [38, 37, 80, 81, 82].

1.2.2 Autoimmune disease

Autoimmune disorders (AIDs) are distinguished by excessive inflammatory response antagonistic towards oneself, resulting in tissue damage or depletion [72, 83, 84]. Based on how they affect the body, AIDs are divided into organ-specific (e.g. multiple sclerosis (MS), type I diabetes (T1D), myasthenia gravis (MG), inflammatory bowel diseases (IBDs), celiac disease (CeD)) and systemic (e.g. systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), Sjögren's syndrome) [77]. AIDs might be driven by autoantibodies or cytotoxic T cells, but in all cases of AIDs, helper T cells are required [77, 78, 85, 86, 87].

A prime example of this is the autoimmune destruction of pancreatic beta cells in type 1 diabetes mellitus [88]. Even though, originating locally in the pancreas its effects are multi-systemic and lethal if left untreated. Signs of both B and T cell hypersensitivity are typically manifested, with the presence of autoantibodies and strong genetic associations with the major histocompatibility complex (MHC) [89, 90].

Autoimmune diseases are complex disorders and are believed to be a result of a combination of genetic and environmental factors (Figure 1)

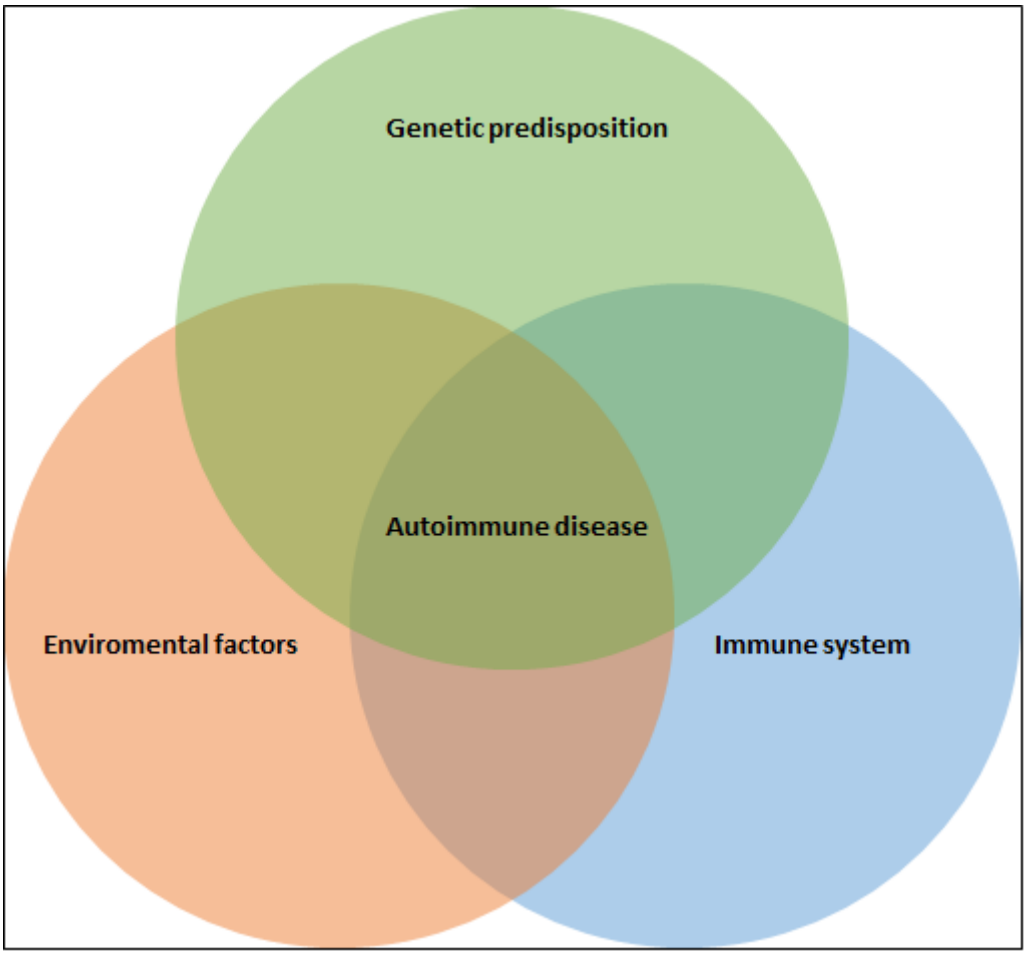


Figure 1. Factors leading to autoimmune disease

The prevalence of autoimmune diseases is approximately 3–5% in the general population and are vastly different in different parts of the world [72]. The prevalence of autoimmune disorders in the western world is thought to be higher than in other parts of the world [91] [72]. The exact number of autoimmune disorders is not known but it is estimated to be around 100 different illnesses [73, 79, 91]. Noteworthy, some autoimmune disorders share certain features and therefore the discrimination between them can sometimes be difficult [92]. Core symptoms for many AIDs often include: low-grade fever, fatigue, malaise, pain and skin changes [83]. Several studies have suggested that ME/CFS shares several of those key features with autoimmune diseases [24]

T cells and autoimmune disease

T cells are crucial in the proper functioning of the immune system. Homeostasis, immune response and immune memory all depend on T cells [78]. T cells express receptors with the ability to recognize various antigens from pathogens, cancers, and toxins, and also maintain

immunological memory and self-recognition and tolerance [93]. T cells are also involved as crucial drivers of many AIDs [78, 94, 95].

TCRs recognize peptides presented by HLA molecules [96]. T cells need to discriminate between self and pathological peptides and are activated upon HLA presentation of foreign epitopes [78, 96]. This recognition, driven by the TCR interaction with HLA is crucial for the immune response, but can also lead to the development of auto-reactivity and autoimmunity [96]. To be able to recognize the highly polymorphic HLA and the large repertoire of peptides presented by it, TCRs also need to be extremely diverse [97, 98].

At the genetic level, there is a complex yet elegant mechanism to generate the vast diversity needed for TCRs ability to recognize antigens. Unlike HLA, TCR derives its diversity from V(D)J recombination [98, 99]. V(D)J recombination is a process of somatic recombination that occurs in developing T and B cells [98]. V(D)J gives rise to an extremely diverse repertoire of antibodies in B cells and TCRs in T cells [100]. There are four TCR loci; *TCRA, TCRB, TCRC* and *TCRD.* Likewise, there are four TCR gene types, variable (V), diversity (D), joining (J) and constant (C) genes. The V domain at the N-terminal end of each TCR chain results from a V(D)J recombination while the constant region is encoded by a C gene [100, 98]. The somatic assembly of those TCR genes produces a diverse T cell range and is an elementary part of thymocyte development. Autoreactive T cells play a key part in autoimmune diseases [101]. They can act both as regulatory and effector cells. Notably, the CD4+ T cell subtype, has been connected with many aspects of autoimmune inflammation [102]. Autoreactive T cells with changed or modified TCR topologies may escape thymic deletion due to changes in binding to the MHC [95]. It is hypothesized that this change in biding pattern enables the cell to avoid apoptosis thus enabling autoimmune responses. [95, 101]

HLA and autoimmune disease

The MHC is called HLA in humans. It is positioned on the short arm of chromosome six and encompasses the HLA class I, II and III genes. The corresponding molecules coded by those genes (class I and II) are amongst the most important immunological molecules in the human body [96, 99].

HLA class I molecules are expressed in all cells except red blood cells and they present epitopes to killer T cells. Killer T cells express cluster of differentiation 8 (CD8) co-receptors and TCR [97]. When a killer T cells CD8 receptor gets in contact with HLA class I molecules,

and the killer T cells TCR matches the epitope on the MHC class I molecule, the T killer cell triggers apoptosis [97]. Therefore, HLA class I helps mediate a primary immunological response to external threats like viruses and bacteria. The classical HLA class I genes comprises of *HLA-A*, *HLA-B*, and *HLA-C* [103].

HLA class II can, given the right circumstances, be expressed by all cell types, even though they are mostly not [104]. Normally they are expressed on antigen-presenting cells (APCs): macrophages, B cells, and dendritic cells. An APC takes up an antigen and provides an epitope which is displayed to the APC's surface coupled within an MHC class II molecule. On the cell's surface, the epitope can be recognized by TCRs. The classical HLA class II genes comprise of *HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4* and *HLA-DRB5*.

Currently (07-2021), there are 30,862 HLA alleles described by the HLA nomenclature and included in the IPD-IMGT/HLA Database [105](Figure2). This vast allelic diversity can possibly be explained by evolutionary pressure to increase the likelihood that at least some can develop an immune response against an emerging infection thereby maximizing the possibility of survival of the species [106]. Two plausible mechanisms, in this case, could be: heterozygote advantages and frequency-dependent selection [103]. Theoretically, if heterozygosity is present at all of the classical class I or II HLA loci, an APC could present over 10^{12} different peptides [106].

The first disease to which genetic variation in HLA was linked was Hodgkin lymphoma in the 1960s [106]. Likewise, HLA loci were the first and strongest associations that were identified with AIDs [107, 108]. Since then it has been established that HLA variation is correlated with a substantial number of pathologies [99]. This correlation with multiple diseases is no surprise, considering that the HLA region is the most polymorphic part of the human genome [99]. The proteins encoded by the HLA genes play crucial roles in antigen processing and presentation as well as inflammatory response. [99]. Therefore, it is to be expected the vast majority of pathologies correlated with HLA are autoimmune and infectious in nature.

Nevertheless, research efforts and the rise of cost-effective and precise genotyping platforms have further illuminated the understanding of the role of HLA associations with AIDs [109]. Studying HLA association in disease is by no means an easy task. When compared to other parts of the genome, the HLA gene complex exhibits high levels of linkage disequilibrium (LD). Due to this LD property of HLA genes, identifying causal and independent genetic

variation has been challenging [109]. The extensive complexity of this genetic region has spawn elaborate computational tools to better study its hypermorphic properties [110, 111]. Regardless of those obstacles, many large-scale studies on HLA and AID associations have been performed in the past. Autoimmune phenotypes like RA [112], CeD [113], T1D [89], SLE [114], PS [115], and MS [116] have been extensively analyzed for their genetic make-up. While all of those illnesses have different mechanisms of action and reflect their autoimmunity features on different levels the HLA association seem to be one feature they all share [117].

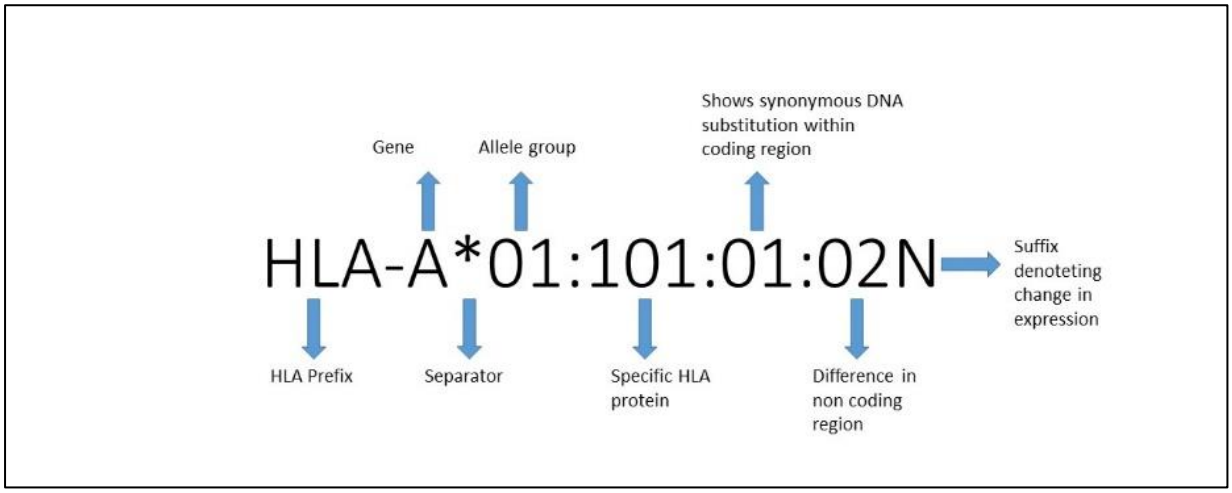


Figure 2. Genetic nomenclature system for HLA alleles

1.3 Genetics and complex disease

1.3.1 Genetic variation in humans

The human genetic composition and regulation are highly complex. The human genome consists of more than 3 billion base pairs, spread on 23 chromosomes, which give rise to around 20000 protein coding genes [118]. About 99.9% of this genetic information is identical in any two given individuals. All human phenotypic variation like, skin and hair colour, height and facial feature lies in the remaining 0.1 % [119]. This diversity is due to a variety of genetic variation, from large structural variations to single nucleotide variants (called SNVs and/or SNPs) [120]. A SNP is described as a change of a single nucleotide that occurs at a specific location in the human genome [120].

The combination of two SNP alleles that a person has on a particular pair of homologous chromosomes is called a genotype [121]. Those genotypes can be either homozygous or heterozygous. Likewise, a haplotype is a sequence of more alleles along a single chromosome

[121]. When multiple associated SNPs, often in LD, in a genetic region are observed, haplotype analysis can uncover causal genetic variation not detected by single SNP analyses [122].

For the most part, complex diseases are caused by an interplay of genetic and environmental factors. A huge portion of all diseases falls into this classification, including most AIDs. It is well established that complex diseases do not adhere to Mendelian inheritance [123]. A person might inherit genetic variants associated with a complex disease but due to the lack of external stimuli will never develop the illness [124]. A good example of such inheritance is celiac disease (CeD). This disease affects the body's ability to degrade dietary gluten thereby triggering an autoimmune reaction which causes a variety of symptoms in a range from bloating to internal bleeding. [125]. Even though findings from the 1970s supported an association with HLA class I alleles later analyses have shown that CeD is strongly correlated with specific HLA class II alleles [108]. This strong HLA class II (HLA-DQ2 and/or -DQ8 haplotype) association provides a good diagnostic tool for the exclusion (negative test) of a CeD diagnosis [126]. A high portion of the general population will carry the HLA-DQ2 and -DQ8 haplotype (>50 %) and therefore the applicability of positive testing is limited [125, 126]. In simple terms, the alleles encoding the HLA-DQ2 and HLA-DQ8 molecules are necessary for CeD development but they are not the single cause of the illness [127].

1.3.2 Linkage disequilibrium (LD)

In studies like GWAS, linkage disequilibrium (LD) is highly relevant since it enables identifying genetic markers that tag the actual causal variants. LD is defined as the non-random association of alleles at two or more loci in a given population. In any large population with no external evolutionary pressure (e.g. selective mating), alleles should be randomly associated with each other and if some alleles are found together more often than we would expect, we conclude that they are in LD (Figure 3).

LD between alleles at two loci can be shown as $\boldsymbol{D_{AB} = p_{AB} - p_A p_B}$ [128]. This represents the difference between the frequency of a pair of alleles at two loci (p_{AB}) and the product of the frequencies of those alleles (p_A and p_B) [128]. Usually, this applies to two genetic variants on the same chromosome, where AB is a haplotype and p_{AB} is the haplotype frequency [128]. D_{AB} should characterize a population but most commonly, D_{AB} is estimated from allele and haplotype frequencies in a given sample e.g. patient population. [128]

In the above-mentioned example, LD is represented with the value *D*. *D* completely shows the range to which two alleles, are non-randomly associated [128]. However, *D* is often not ideal when comparing LD at different pairs of loci because of the range of possible D values [128]. Because of that LD is most commonly represented by D' and r^2 [129]. D' is defined to be the ratio of *D* to its highest possible absolute value, given the allele frequencies [128, 129]. D' has an advantage over D because when $D' = 1$ it suggests that at least one of the four possible haplotypes is absent, regardless of the allele frequencies [128]. This is described as a 'perfect/complete' LD [128, 129]. r^2 on the other hand can be defined as the square of the statistical correlation between two alleles [128]. When $r^2=1$, two alleles are in perfect LD i.e. the two analyzed alleles will always be inherited together in a given population [130]. By and large, r^2 and D' are similar in that both can be nearly one even if one or both alleles are in low frequency [128]. Nevertheless, *D'* is a one-directional i.e. high D' does not mean that if we know the allele at one locus that we can predict the allele at the other locus with high accuracy. On the other hand, r^2 is a bidirectional LD measure i.e. if we know the allele at one locus we can predict, with a high degree of accuracy, the allele at the second locus. This difference makes r^2 more suited for SNP imputation then *D'*

Understanding the LD pattern can improve our understanding of the genetic basis of human complex traits like AIDs. It gives the possibility for more efficient fine-mapping strategies of genes directly involved in many pathologies [131].

In the case of AIDs, the most interesting genetic region to look into LD structures is certainly the HLA loci with the high degree of LD that exist between alleles at neighbouring loci [131].

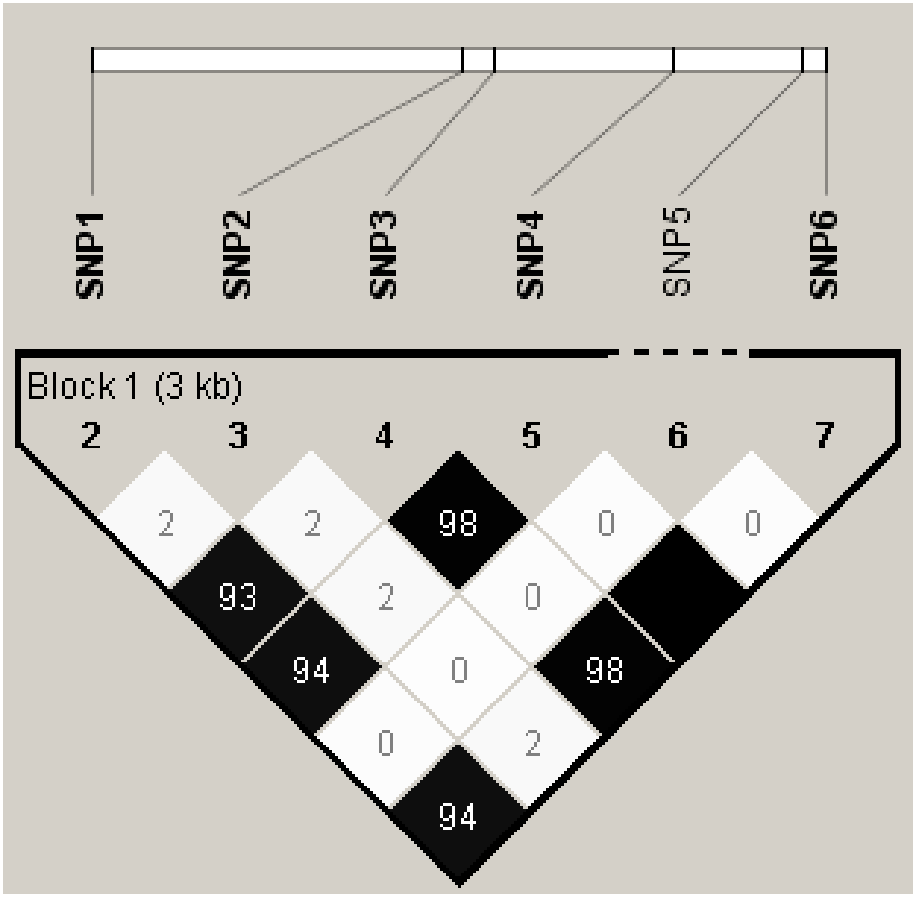


Figure 3. The LD between 6 SNPs measured in r^2 shows how well these SNPs are correlated. Values shown are 100 × r^2.

1.3.3 Genome wide association studies (GWAS)

A GWAS is any study on a genome-wide set of genetic variants in a population to inspect if any of those variants are associated with a trait or phenotype [132] (Figure 4). They are conducted to establish a correlation between certain genetic variants and a phenotype/disease. The frequency in which a particular variant occurs is compared between a control and disease set of participants. However, GWAS analysis have noteworthy limitations, which have to be addressed properly to extract biologically relevant information [133, 134, 135]. Many diseases are characterized by a rather small number of loci with large effect sizes [134]. Other diseases may have a more complex genetic composition with a large number of associated loci with lower effect sizes. Broadly speaking, a disease is either controlled by rare variants (e.g. muscular dystrophy, cystic fibrosis, or Huntington's disease), each having a large effect on the phenotype, or many common variants with only a small overall phenotypic effect as is typical in so-called complex diseases (e.g. diabetes, hypertension, schizophrenia, or autism) [136].

An interesting finding of the first GWAS was that even preeminent loci in the human genome have minor effect sizes and that the significant associations only explain a small amount of the predicted genetic variance, hence, coining the term “missing heritability” [137]. This has been largely solved by analyses showing that common SNPs with p values below genome-wide statistical significance account for most of the “missing heritability” of many phenotypes/diseases[137]. Several studies have demonstrated that significant genetic variants are present in active chromatin sequences (e.g. promoters and enhancers) and in the case of AIDs, they cluster in active chromatin sequences of B and T cells [135, 138, 139, 140]. Even though those associated loci are assumed to be pivotal in gene expression regulation, it is not clear in which physiological contexts all of this occurs. This has hindered the translation of GWAS findings into clinical interventions [135, 137].

GWAS have successfully identified several hundred of loci associated with AIDs [139, 140, 141]. Some AIDs like RA [142], CeD [143] and T1D [144] have several dozens of established risk loci.

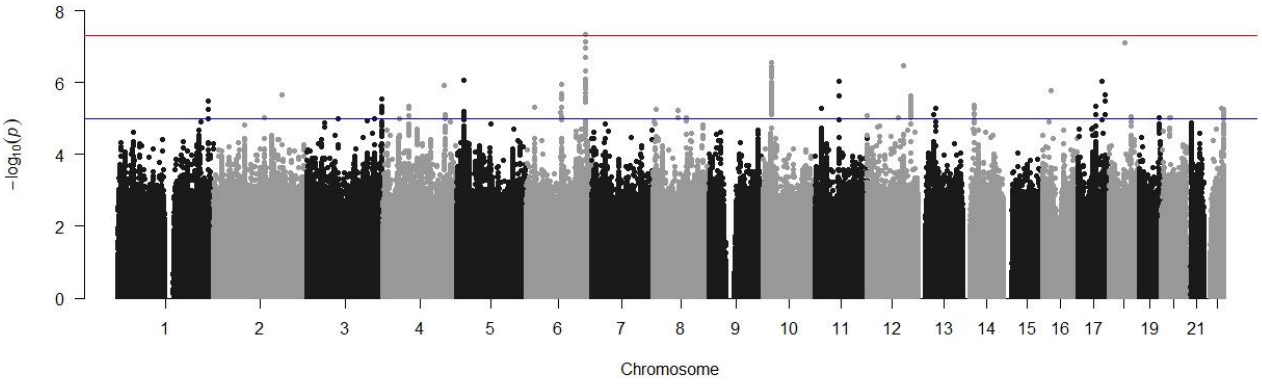


Figure 4. Typical Manhattan plot from a GWAS with -log10 transformed p values on the y axis showing the significance levels of each genetic variant (dots) tested for association with a phenotype. Chromosomal locations on the x axis

1.3.4 Shared genetics in autoimmune disease (AID)

For the majority of instances, a particular AID manifests without previous family occurrences [145]. Nevertheless, aggregation of diverse AIDs in families is well reported [145, 146]. Evidence that genetics plays a crucial role in disease development has been provided by familial occurrences and linkage studies [72, 106]. In contrast to monogenic diseases, complex diseases such as many AIDs are associated with multiple genetic variants, conferring varying, mostly small, effects on disease susceptibility and development [123].

While the familiar clustering approaches with linkage studies had given some insight into disease susceptibility, the application of GWAS to studies of AIDs has expanded the understanding of these disorders with an enormous number of disease-associated loci [141, 147].

Even though genetic studies in AIDs have reported a multitude of loci and markers, non are approaching HLA significance [109]. Most associations identified in AIDs are with HLA class I and/or class II genes [141].

For many AIDs where pathology is characterized by autoantibodies, the most significant associations are with HLA class II genes. Those are diseases like SLE, CeD, RA or T1D [113] [114]. Where the autoantibody symptomatic is not the case , like in PS or ankylosing spondylitis, those illnesses tend to have a stronger HLA class I association [115, 148, 149] [150].

Besides the shared HLA associations AIDs share also a substantial amount of non-MHC associated loci [151](Table 2). A huge majority of these genetic variants are proximal genes or intergenic variants regulating genes involved in the immune system [152, 153, 154, 155]. Those genes are often expressed in immune cells, like B and T cells, which are the main mediators of autoimmune disease. For example, genes like *PTPN22* or *STAT4* are found to be associated with many AIDs [141]. Because of shared loci in different AIDs, combinations of genetic factors may determine the disease of any particular patient [141]. In genes like *PTPN22* and *IFIH1* the effect of the same genetic variant is antipodal in different AIDs [141]. Furthermore, the associated variants can differ in different AIDs, possibly pointing to the presence of multiple disease-causing variants in the locus [141].

Table 2. Some of common risk loci shared between some AIDs

Autoimmune diseases compared	**Shared risk loci**
T1D and CeD [156]	*RGS1, IL18RAP, TAGAP, CCR3, PTPN2, CTLA4, SH2B3*
RA and CeD [157]	*SH2B3, TNFAIP3, IL2/IL21, SH2B3, LPP, MMEL1/TNFSF14, PF KB3/PRKCQ*
CD and CeD[158]	*PTPN2, IL18RAP, TAGAP, PUS10*
SLE and 16 AIDs [159]	*IL23R, OLIG3/TNFAIP3, and IL2RA (most shared between AIDs analyzed)*
T1D and other AIDs[160]	65% of 112 identified risk loci are common with other AIDs *(e.g PTPN2, BACH2, CTLA4, CD226, UBASH3A)*

2 Thesis aims

Our project is based on the hypothesis that autoimmunity or immune dysregulation is involved in the pathogenesis of ME/CFS.

Our primary aim was to conduct a thorough genetic screen of immunological genes in a large ME/CFS patient group diagnosed according to the Canadian consensus criteria.

In addition to this, we aimed to fine-map the HLA region in an effort to see the primary association driving elements in the region.

In order to fulfil those aims we:

- Fine-mapped the previously observed HLA associations in ME/CFS by adding five more HLA loci and 5,342 SNPs (**Paper I).**
- Performed a GWAS like immunochip (iChip) study in ME/CFS, using a large Norwegian ME/CFS patient group (diagnosed according to the Canadian Consensus, Criteria). Additionally, we examined a Danish replication cohort (diagnosed according to the Canadian Consensus Criteria) and the UK Biobank CFS population. (**Paper II)**
- Finally, we wanted to see whether TCR genetics was associated with ME/CFS (**Paper III**).

3 Summary of papers

Paper I

Human leukocyte antigen (HLA) associations are one of the most robust features of immune-mediated and autoimmune diseases. Previous HLA studies have been performed in ME/CFS prior to our work, however, without common results. Additionally, those studies were substantially underpowered and used varying diagnostic criteria to assess ME/CFS. Hence, our main objective in **paper I** was to expand previously performed work by our group which reported associations between ME/CFS and *HLA-C:07:04* and *HLA-DQB1:03:03* alleles. However, the HLA region consists of numerous immunologically relevant, and also manifests extensive and complex linkage disequilibrium (LD) across the region. We wanted to further investigate the associations observed in the HLA region by genotyping five additional classical HLA loci and < 5000 SNPs in 427 Norwegian ME/CFS patients, diagnosed according to the CCC, and 480 healthy ethnically matched controls. Analysis revealed two separate association signals in the HLA class I and II regions ($P \leq 0.001$). Moreover, the main association signal in the HLA class II region was located within the HLA-DQ gene region, most likely due to *HLA-DQB1*, particularly the amino acid position 57 in the peptide binding groove, or an intergenic SNP upstream of *HLA-DQB1*. In the HLA class I region, the signal was less clear and could possibly map outside the classical HLA genes. Interestingly, we also observed that almost 62% of responders to cyclophosphamide treatment for ME/CFS carried either the rs4711249 risk allele and/or DQB1*03:03 vs 12% of patients not responding to the treatment. Our findings suggest the involvement of the HLA region, and in particular the *HLA-DQB1* gene, in ME/CFS. These findings need to be replicated in much larger cohorts considering our relatively small sample size. In particular, investigation of the involvement of *HLA-DQB1*, a gene the largest risk for well-established autoimmune diseases, is warranted

Paper II

The presence of hereditary components in ME/CFS is supported by excess relatedness. However, GWAS in ME/CFS have been limited by small sample sizes and broad diagnostic criteria, therefore no established risk loci exist to date. In **paper II** we aimed to investigate immunologically relevant genes using the iChip. In addition, we aimed to improve on previous study designs in the filed by increasing overall power and using clinically well-characterized patient populations. We analyzed three ME/CFS cohorts: a Norwegian discovery cohort of 427 ME/CFS patients (CCC), a Danish replication cohort of 460 ME/CFS patients (CCC) and a replication dataset from the UK biobank of self-reported CFS patients 2105. To the best of our knowledge, this is the first ME/CFS genetic association study of this size including more than 2900 patients (887 CCC). Even so, we did not find any ME/CFS risk loci displaying genome-wide significance, however, we identified six regions encompassing the *TPPP*, *LINC00333,RIN3*. *IGFBP1/IGFBP3*, *IZUMO1/MAMSTR* and *ZBTB46/STMN3* genes that have indicative associations. Interestingly, these genes are expressed in disease relevant tissues, e.g. brain, nerve, skeletal muscle and blood, including immune cells like T cell subsets, B cells, NK cells and monocytes. Furthermore, several of the ME/CFS associated SNP genotypes are associated with differential expression levels of these genes. Nevertheless, considering our relatively small cohort size as well as low allele frequencies for most of our top hits our reported results need to be further investigated in much larger cohorts. However, considering that our study is the first of its kind in size and cohort characterization, we believe our generated datasets represent an important stepping stone for the future understanding of the genetic influence on ME/CFS development.

Paper III

An increasing number of studies have implicated the involvement of several different parts of the immune system in ME/CFS. The role of autoreactive T cells in established AIDs is well documented. However, the role of TCR genetic polymorphisms is still unclear. TCRs do not only derive their diversity from polymorphisms in TCR loci but rather from somatic recombination. This is in stark opposite to HLA molecules which derive their diversity by their unique and extensive polymorphicity in their corresponding HLA genes. The main aim of **paper III** was to investigate previous claims by a GWAS in 42 ME/CFS patients which reported significant association with two SNPs in the T cell receptor alpha (TRA) locus ($P < 5x10-8$). Additionally, our observed results from **paper I** warranted investigation of the TRA locus. We wanted to replicate the previously reported findings in the TRA locus using a large Norwegian ME/CFS cohort (409 cases and 810 controls) and data from the UK biobank (2105 cases and 4786 controls). We investigated numerous SNPs in the TRA locus, including the two previously ME/CFS associated variants, rs11157573 and rs17255510. No associations were observed in the Norwegian cohort, and there were no significant association with the two previously reported SNPs in any of the cohorts. Albeit, other SNPs showed signs of association ($P < 0.05$) in the UK biobank cohort and meta-analyses of Norwegian and UK biobank cohorts, none survived correction for multiple testing. Hence, we did not identify any reliable associations with variants in the TRA locus.

4 Methodological considerations

4.1 Study population

4.1.1 Diagnostic criteria

A key factor in genetic studies like ours is accurate and standardized phenotypic discrimination [161]. ME/CFS has a wide range of definitions and is clinically heterogeneous. The lack of objective and reliable biomarkers makes it hard to establish a firm homogeneous patient population. In order to minimize heterogeneity, the CCC were chosen for the Norwegian and Danish populations. This set of diagnostic criteria has a more standardized approach in comparison to other widely used criteria like the Fukuda criteria [1, 11]. However, even with the use of more strict diagnostic criteria like the CCC the patient heterogeneity may be substantial [10]. This is in part due to variation in the evaluation of the patients by the physicians making the initial diagnosis. The mandatory symptoms and exclusionary diagnosis are often hard to assess for physicians not specializing in patients with ME/CFS [162]. The assessment of malaise, pain or even psychiatric disorders is to a high degree subjective and prone to error and misdiagnosis. Moreover, whether the CCC truly represent a more homogeneous population in comparison to other diagnostic criteria is not certain [10]

Additionally, the UK Biobank population was comprised of self-reported CFS patients which increased the heterogeneity of our overall patient population in **paper II.** It is fair to assume that some percentage of patients from the UK Biobank fit the CCC however the exact number is undeterminable**.**

4.1.2 Genetic background of patients

Another important aspect of our study population is the genetic background. Besides phenotypic homogeneity for the patients, ethnic/genetic homogeneity plays a crucial role in genetic epidemiological studies like ours [161, 163]. The Norwegian population only included ethnic Norwegians (self-reported ancestry). However, ethnic information was lacking for the Danish population. In order to minimize the effects of ethnic diversity in the UK Biobank we included only patients which self-identify as “British,” “Irish,” “Any other White background” and “White,” In addition to the exclusion of patients in accordance to their self-reported heritage, we also performed a principal component analysis of the Norwegian and UK Biobank datasets. A principal component analysis is a standard method to correct for population stratification and is used to cluster individuals by ancestry when using genome-wide data [164]. This is made possible by comparing the genetic variation of our group of

interest (the Norwegian and UK populations) to the 1000 genomes reference data. We did not observe divergent ancestry in our principal component analysis hence we concluded that our given populations are sufficiently genetically homogenous. This approach, however, could not be used for the Danish datasets since we had only 20 SNPs available which are far below the thousands of SNPs needed across all chromosomes to conduct such an analysis.

4.1.3 Control data

The control data from the Norwegian and Danish populations were previously used in other studies [165, 166, 167] while the UK Biobank controls were randomly chosen from the non-CFS population of the UK Biobank. Some of the controls could develop ME/CFS after the point of inclusion as controls and therefore decrease the studies power to detect true differences. This is especially true for the UK biobank controls given the poor phenotyping discrimination to begin with. However, the effect of future ME/CFS development in the Norwegian and Danish controls can be assumed to be relatively small given the general prevalence of ME/CFS (<0.5%).

4.2 Genotyping methods

4.2.1 Microarray-based genotyping

For **all three papers,** the iChip array was used to genotype SNPs covering regions previously associated with other diseases. While this approach is more targeted than an ordinary GWAS array it comes with its drawbacks [168]. One weakness of this targeted approach is that the iChip does not cover the whole genome (only approx. 250,000 SNPs), and depends on the power of the previous GWAS studies for its markers [168]. It is possible to argue that in a poorly researched disease like ME/CFS it would be more indicative to use whole-genome chip which covers millions of SNPs. However, the evidence of immune involvement in ME/CFS from previous studies warranted the use of this more targeted approach. The iChip also contains a dense set of SNPs in the MHC and KIR/LILR regions which allowed for imputation. The Norwegian (**paper I, II and III**) and Danish (**paper II**) controls were genotyped using an older version of the iChip hence enabling us only to use the overlap between iChip versions 1 and 2 (approx. 150,000 SNPs).

The UK biobank data were genotyped using the UK Biobank Axiom Array (Thermofisher). Even though different genotyping arrays were used, recent studies have shown that the inter platform reliability of obtained genotypes lies around 99% [169]

4.2.2 Open Array Taqman and Taqman Assays

We have also used Open Array Taqman (QuantStudio 12K Flex Real-Time PCR System) for genotyping of the 24 tag SNPs in the Danish cohort in **paper II.** Four SNPs were excluded from analyses due to poor genotyping and QC. Likewise, we only selected tag SNPs from the directly genotyped SNPs and not the imputed SNPs which showed stronger association. For **paper III** we also used single Taqman assays to genotype additional SNPs in the TRA region.

4.3 Genotype imputation

Genotype imputation is a process of estimating missing genotypes from a reference panel [170]. Imputation boosts the power of detecting SNPs in GWAS. In **paper II** we imputed additional genotypes for our Norwegian cohort and increased the total amount of available genotypes from < 150,000 to > 1,500,000. The performance of genotype imputation can be affected by factors like the particular imputation software used, sample size, reference panel used and input SNP density [170]. For our imputation, we used the Michigan imputation server, which uses the Minimac 4 imputation tool (Reference Panel: 1000G Phase 3v5 EUR, rsq filter R<0.3, phasing via Eagle v2.4, Build 37) [171]. While studies show generally excellent imputation results across several different software and reference panels, lower frequency alleles (MAF < 5%) tend to have a lower imputation accuracy rate [170, 172, 173]. This consideration applies to some of our association signals in **paper II** since several of them had minor allele frequencies below 5 %. Additionally, the UK Biobank imputed genotypes had been calculated using the IMPUTE2 tool. However, as mentioned before imputation results across different software options are reliable [170, 172, 173].

4.4 Quality control of the data

For **all three papers,** the SNP data was stringently quality controlled. This is very important in genetic studies similar to ours. The lack of stringent quality control can lead to significant changes in the results and thus lead to either false positive or negative results [163, 174]. For all three datasets (Norwegian, Danish, UK Biobank), we excluded SNPs below a 95% genotyping success rate as well as SNPs having an allele frequency (MAF) below 1%. Likewise, we excluded SNPs deviating from Hardy Weinberg equilibrium (P= 0.001). In addition to this, in the Norwegian dataset, we manually inspected genotyping cluster-plots from iChip v2 and poor performing SNPs were excluded prior to further analyses. However,

this was not done by us for the control data in either cohort or the UK biobank data since the data was obtained without genotyping cluster plots.

4.5 HLA typing and imputation

Due to high degree of polymorphisms in the HLA region, sequencing is challenging when compared to other genomic regions [175]. Several methods for HLA typing like sequence-specific oligonucleotide (SSO), sequence-specific primers (SSP) and Sanger sequencing are available [176]. However, those approaches are low throughput and time consuming [177]. That is why we opted for HLA genotyping by next-generation sequencing (NGS) which is more time and cost-effective as well as high throughput [175].

In **paper I**, high-resolution HLA genotyping by NGS was performed at the Department of Medical Genetics, Oslo University Hospital for ME/CFS patients, while HLA genotyping for the 480 healthy controls was performed by NGS at Stanford University, USA [178]. However, samples were missing a few HLA genotypes which is why we opted to impute the absent ones from the available SNP data obtained via the iChip (HLA*IMP:03, University of Melbourne). To ensure that the imputation was reliable and did not deviate from the real genotypes we also imputed HLA genotypes for all samples not missing any HLA genotypes in order to compare the outcome of the imputation to the observed genotypes. A concordance > 99 % was observed, hence enabling us to take the imputed HLA genotypes as true.

4.6 Meta-analysis

Sufficient statistical power is of crucial importance in finding new genetic associations. A meta-analysis can increase the power to detect association signals. In **paper II,** we combined multiple data sets to do a meta-analysis of the individual findings. There are several crucial steps to ensure a properly performed meta-analysis [179]. To conduct a proper meta-analysis, all datasets should be as homogeneous as possible and represent the same phenotype [179]. This is, however, hard to ensure with ME/CFS given the wide range of diagnostic criteria.

Still, a meta-analysis method is very useful to find out if findings are consistent and if they can be applied to different populations or identify interesting patterns among studies [179-181].

An important consideration in meta-analysis is whether to use a fixed or a random-effects size model [180]. In a fixed-effects size model, the assumption is that all of the studies have a

common genetic effect and that the study-specific findings only differ from one another because of sampling variation. Contrary to this, in a random-effects size model, it is assumed that each study population has its own size of genetic effect and that the goal is to estimate the average effect overall potential populations. It is not easy to choose between those two models especially in the context of ME/CFS. The choosing of either of the two can have a major impact on the P-values. When the inappropriate model is used the P-values can be poorly adjusted i.e they will not achieve their nominal type-one errors [179]. A fixed-effects size model applied in a heterogeneous data set will tend to overestimate the P-value[179]. However, random-effects meta-analysis could be too conservative in their P-value estimation [179]. Additionally, it is possible that some SNPs will have the same effect in all populations, while other variants have effects that vary across populations [180].

4.7 Statistical considerations

The power of the experiment can be defined as the ability to identify genetic variants that truly represent phenotypic variations [163]

Due to the heterogeneity of our ME/CFS phenotype and likely multifactorial aetiology, genetic risk variants can be assumed to have rather moderate or small effect sizes. Hence, enough statistical power is needed to detect those variants. In the case of GWAS a beta of 0.8 (80%) is considered good [163]. When doing a power calculation for **paper II** (427 cases, 972 controls, significant level $p < 5x10^{-8}$ and disease prevalence of 0.5%) we get that our study has a power of 0.13 (13%). In GWAS, determining the right P-value cutoff for statistical significance is crucial to minimize the number of false-positive and false-negative associations. There are several statistical methods that try to account for multiple testing [182]. The Bonferroni correction, Sidak correction, False Discovery Rate (FDR), permutation test, and Bayesian approaches are all ways of adjusting for multiple testing [182]. The Bonferroni correction is considered the most conservative way of setting a P-value threshold since it assumes independence of the genetic variance from each other [182].The genetic variants investigated are dependent on population-specific elements, like LD and minor allele frequency [182]. Hence, the set appropriate threshold for genome-wide significance might vary in different populations [182].

Since no classical GWAS was performed, in **paper II,** but rather a targeted GWAS-like approach with the iChip was undertaken one might argue that a different threshold for

significance could be set. Nevertheless, the GWAS significance threshold ($P < 5x10^{-8}$) remains the gold standard in detecting true associations in studies like ours [183]. The significance threshold (P<0.001) for **paper I** was set primarily due to the previous findings it was leaning on to [63]. We used the same ME/CFS cases but had almost 1/10 of the controls (480 vs 4552).

All three papers used in this thesis were underpowered to detect GWAS level significant associations. However, our work represents by far the largest genetic analysis in ME/CFS to date and it can be considered a first step in establishing larger consortia type efforts.

4.8 Ethical considerations

In this project, we followed all ethical and scientific standards in accordance with the Norwegian ethical guidelines. Our studies are approved by the Regional Committees for medical and health research Ethics (REK number 2015/1547)

The ME/CFS patient population is diverse both in background and in clinical manifestation. One of the core symptoms of ME/CFS, PEM, can possibly affect patients ability to fill out questionnaires related to ME/CFS diagnosis. Likewise, blood sampling can also cause distress in the most severe patients. It is therefore advised to conduct those mentioned actions with caution and maximal care in order to not produce extra suffering.

Additionally, the importance of a clear explanation of possible side effects with treatments like cyclophosphamide is of utmost importance as for the data included in **paper I** from the study by Rekeland et al. [184]. Common side effects of cyclophosphamide treatment include constipation, nausea, abdominal pain, hair loss, changes in skin colours with some patients experiencing severe side effects like chest pain, difficulty breathing or swallowing and bleeding [184, 185]. Research using drugs like cyclophosphamide which can possibly increase patient suffering always need to be precisely planed and monitored.

It is important to add that patients are very active and supportive of research. Due to their impacted life quality, every scientific finding is welcomed with open arms. However, this enthusiasm can also lead to misinformation being passed around in the community. This is why patient outreach and clear, unequivocal communication with patients is of great importance. Scientist researching ME/CFS should always keep in mind that, due to the suffering caused by the disease, some patients expectations from individual studies are too high. In the case of our work, patients might want to obtain their HLA type or risk alleles.

Hence, it needs to be clearly communicated how research is conducted and what the results in the individual studies mean.

All human research tends to exhibit more or less ethical considerations, but given the suffering ME/CFS causes and the knowledge gap about the disease in general those considerations are, in our opinion, well balanced in this work.

5 Discussion

The work in this thesis has primarily focused on investigating genetic regions with known immunological correlation in autoimmune disease. The main focus was to apply established ways of genetically researching autoimmune disorders onto ME/CFS in order to investigate the immunogenetic component of the disease. It has been shown that common genetic risk factors exist between autoimmune disorders [109]. Furthermore, previous immunological studies have reported that ME/CFS shares features with established autoimmune disorders [24], with also some studies suggesting significant associations with genes previously reported to be associated with autoimmune disease [56, 60]. Considering our hypothesis that ME/CFS is, at least in part, immune-mediated those findings come as no surprise taking into account substantial genetic and clinical overlap between autoimmune disorders. Thus, investigating common autoimmune susceptibility loci is warranted and can possibly shed light on the genetic structures of ME/CFS. In **all papers** comprising this work, investigating the same loci and genes previously associated with autoimmune disorders was, therefore, a well-considered approach. Additionally, in **paper I** the extensive LD in the HLA region complicates the pinpointing of primary risk factors of previously reported associations by our group, hence, a fine-mapping approach was employed.

5.1 ME/CFS as an autoimmune disease

Due to the clinical nature of ME/CFS, replication of research results has been harder than in many other diseases. The phenotype of ME/CFS is less clearly defined, and in contrast to several established autoimmune disorders, ME/CFS lacks biomarkers and/or distinct clinical features which could indicate the target tissue or organ. Villous atrophy in CeD [186], skin changes in PS [115] or joint inflammation in RA [112] are all ideal indicators of pathological inflammatory process. In diseases like seropositive RA (anti-cyclic citrullinated peptides), SLE (antinuclear antibodies) and celiac disease (tissue transglutaminase antibody) enhanced levels of auto-antibodies are detectable and serve as reliable diagnostic tools [187, 188, 189]. However, in autoimmune diseases of the central nervous system (e.g. MS or Rasmussen's encephalitis) and IBD, no indicators of humoral autoimmunity are present. Nevertheless, regardless of the mediators of autoimmunity, extremely strong associations with the classical HLA class I and II genes are present across almost all autoimmune diseases.

The research field on ME/CFS is in its infancy compared to other autoimmune diseases, as evident by the number of publications on PubMed (N=578 on 20th of November 2021) compared MS (98,432) or RA (144,480). Yet, while ME/CFS still lacks classical autoimmunity biomarkers and associations, a multitude of studies, investigating several aspects of the immune system, have pointed towards the involvement of the immune system in disease development [24, 30, 33, 34, 35, 36, 37, 38, 40, 41, 42]. Nevertheless, many of those studies have conflicting results and a low level of reproducibility. There are conflicting reports on autoantibody presence, cytokine activity and infectious triggers in the disease [30, 34, 35, 36, 37, 38]. However, several studies have pointed towards NK cell alterations in ME/CFS [33]. Most notably impaired NK cell cytotoxicity seems to be a reoccurring finding in multiple studies[33]. While most of the studies are limited by small and heterogeneous patient populations, it is still indicative that they all seem to point towards the immune system as being a key part of ME/CFS.

Considering that a multitude of studies has shown overlap between ME/CFS and established autoimmune features, analyzing HLA genes and other genes associated previously with AIDs was warranted in our study.

5.1.2 HLA associations in ME/CFS

The primary research question in **paper I** was to fine map the previous HLA association identified by Lande et al. [63]. The background for the initial work by our group is that HLA genes play crucial roles in AIDs. Many individual HLA alleles are associated with immune-mediated disorders. Hence, given the hypothesis that ME/CFS is immune-mediated, it was a natural choice for scientific investigation. Using a case-control study consisting of 426 ME/CFS patients and 4521 controls, Lande et al. observed significant associations in *HLA-C*07:04* and *HLA-DQB1*03:03* alleles [63]. To further pinpoint the primary association signals, we expanded on the previous study by adding five extra HLA loci and 5342 SNPs from the MHC region. Furthermore, this enabled us to also investigate if those findings were two distinct association signals or possibly due to the complex LD in the region. It is important to add that while we used almost identical numbers of ME/CFS cases, we used less controls (480 vs 4521).

The main finding in **paper I** was that the two association signals in the *HLA-C* and *-DQB1* loci are independent of each other. In addition, the associated SNPs were dependent on either the *HLA-C* or *-DQB1* locus. Amino acid residue analysis revealed that HLA-DQB1 residue 57D was positively associated with ME/CFS, but also SNPs upstream of *HLA-DQB1* which

could have a regulatory role. While we could not observe the same significance levels as Lande et al. [63], due to lower statistical power, we observed similar effect sizes for the *HLA-C*07:04* and *DQB1*03:03* alleles. Intriguingly, HLA-DQB1 amino acid residue 57D is coded by *DQB1*03:03* and several other alleles.

We initially intended to also analyze killer-cell immunoglobulin-like receptors (KIR) genes for this study. KIRs, are a family of type I transmembrane glycoproteins expressed on the plasma membrane of natural killer (NK) cells and some T cells [190]. KIR regulate the apoptotic function of NK and T- cells by interacting with HLA class I molecules [96] and is reported to play a predisposing role in many autoimmune diseases [191]. A previous study investigated KIR genes and their HLA ligands in a CFS patient cohort and reported that the presence of *KIR3DS1* may confer susceptibility to CFS [192]. However, the study was small (n=46) and diagnostic criteria are not stated [192]. Unfortunately, due to time constraints, we were not able to investigate KIR-HLA interactions in regards to ME/CFS. However, KIR ligand defined subtypes of the HLA class I alleles were analyzed with no significant associations observed (unpublished data).

When comparing our results to other HLA associations in established autoimmune disorders, we see that the strength of our associations (OR=1.6) is below that what was observed in statistically similarly representative cohorts in other diseases. A Japanese study including 410 classical autoimmune T1D and 257 patients with fulminant T1D observed associations in the HLA-DR region with an effect size of OR = 3.18 and a significance reaching $P = 1.56 \times 10^{-23}$ [193]. Similarly, a study in UC (n=748) and CD (n=979) reached equally significant associations with OR= 4.44, $P = 1.6 \times 10^{-70}$ [194]. Even studies with substantially lower amounts of patients have found GWAS significance levels of HLA associations in diseases like neuromyelitis optica [195]. However T1D, neuromyelitis optica, ulcerative colitis and CD are diseases with unique clinical features which enable well-characterized cohorts. In the case of ME/CFS firm clinical discrimination is lacking hence diversifying the study population. Hence, the ME/CFS phenotype might be more clinically heterogeneous lowering the effect size when studying the current. However, it should be noted that the HLA relative risk ratios in autoimmune diseases range from <2 to >150 [196]. Because of this, large consortia recruited cohorts with thousands of clinically well-characterized patients are needed to establish subgroups of patients.

The ability of HLA molecules to influence thymic selection as well as peripheral anergy of T cells makes its role in the development of autoimmune disease clear. However, the contribution of specific HLA loci or alleles in autoimmune disease is poorly understood in many cases. Celiac disease is a unique example of a disease where both the immunologically involved gliadin peptides from gluten, as well as autoantibodies have been discovered [197].

Even though HLA associations are identified, like for ME/CFS in our **paper I**, they are usually not suitable for being used as biomarkers due to low sensitivity and specificity. However, for some diseases, they are in fact used diagnostically, i.e. CeD, spondyloarthritis and narcolepsy. In spondyloarthritis, testing for HLA-B27 has become a routine clinical test even without a clear understanding on how HLA-B27 contributes to pathogenesis [198]. Likewise, the involvement of HLA-B51 in Behçet's Disease remains unclear albeit up to 75% of patients are carriers compared to only 10% of the healthy population [198]. In RA, *HLA-DRB1* testing is neither included in classification criteria nor recommended as a diagnostic tool by ACR/EULAR guidelines regardless of the strong genetic association [187]. However, *HLA-DRB1* variants could possibly give important insight in regards to the choice of treatment for rheumatoid arthritis [199, 200]. In contrast, the specific HLA allelic make-up is a strong determinant of gliadin presentation and disease risk in CeD. Homozygous CeD patients tend to have a more severe disease manifestation [201, 202]. HLA-DQ2.5 homozygotes exhibit a severe immunological reaction following gastrointestinal infections because of increased interferon-gamma concentrations, which, in hand, has been shown to influence HLA-DQ expression [198, 201, 202]. Different HLA haplotypes recognize specific ligands with specific affinity, hence giving differing risk profiles for celiac disease. When comparing DQ2.5-ligands with DQ2.2-ligands, it is shown that DQ2.5 can present a wider range of gliadin peptides than DQ2.2 [201, 202]. Gliadin cannot be completely digested by human digestive enzymes, hence, DQ2-molecules can recognize these resistant epitopes and present them to CD4-positive T-lymphocytes [201, 202, 203]. DQ2 affinity to gliadin is additionally enhanced by the enzyme tissue transglutaminase (tTG), which changes glutamine residues of gliadin. Glutamic acid, which is the result tTG interacting with glutamine, has an increased affinity to DQ2 molecules [198, 201, 202, 203]. The initialization of autoimmune responses against tTG, TNF alpha and interferon-gamma secretion, ultimately result in tissue degradation and disease development [198, 201, 202, 203].

Our observed HLA associations from **paper I** warrant future replication efforts in statistically more representative, well-defined cohorts. In particular, the putative involvement of *HLA-*

DQB1, a gene important for antigen-presentation to T cells and known to harbour alleles providing the largest risk for well–established autoimmune diseases, needs to be established firmly.

While HLA associations represent the main overlap between AIDs, several other autoimmune specific regions have been identified with the advance of GWAS. Therefore, we investigated those regions using the immunochip (iChip) which was specifically designed for undertakings like ours.

5.1.3 Immunogenetic screening of ME/CFS

The iChip has proven to be a useful tool in identifying novel and replicating existing non-HLA associations in diseases like CeD, RA and T1D [204, 205]. In addition, GWAS/iChip studies have demonstrated that many genetic regions are associated with multiple autoimmune diseases, hence suggesting that autoimmune disorders are likely to share molecular mechanisms of disease pathogenesis [139, 151, 154, 205].

The primary research rationale in **Paper II** was to conduct a targeted GWAS with the iChip. Since our hypothesis was that ME/CFS is an immune-mediated disease, we wanted to investigate previously associated risk SNPs in other autoimmune disorders.

In paper II, we did not identify any consistent and GWAS significant associations. The initial discovery phase, using the Norwegian ME/CFS patients, revealed several regions reaching suggestive association levels of $P<1x10^{-5}$. Most notably, several SNPs in the intronic part of the *TPPP* gene were most strongly associated with ME/CFS. Nevertheless, we could not observe associations of the selected *TPPP* SNPs in the Danish replication cohort. Our replication strategy was, however, suboptimal and the ability of our selected tag SNPs to capture the association was limited. Hence, a whole-genome analysis of the Danish ME/CFS patients could give more insight into the putative *TPPP* involvement. Possibly interesting, however, is that the *TPPP* SNPs, identified in the Norwegian discovery cohort, also showed associations in the UK Biobank cohort. The *TPPP* gene, which codes for the tubulin polymerization promoting protein, plays a key role in the myelination of oligodendrocytes [206]. The degradation of myelin is a key event in MS, hence the reported involvement of *TPPP* in shortened disease duration in MS should come as no surprise [207].

In addition to *TPPP*, several other genetic regions showed associations in our combined analysis. One of these being SNPs near *RIN3*. Recent studies have shown *RIN3* to be associated with Alzheimer's disease [208]. This could be interesting since evidence suggests

that Alzheimer's disease is not limited to the neuronal compartment, but also strongly interacts with the immune system and possibly causing neuroinflammation [209]. Both *TPPP* and *RIN3* are associated with neuro inflammatory states in MS and Alzheimer's disease, maybe pointing towards the immune microenvironment of the brain as a possibly involved in ME/CFS. This speculation is strengthen by studies indicating that neuroinflammation was present in several brain areas of patients with ME/CFS [210, 211, 212].

The associations we observed are worthy of follow up; however, an important issue with **paper II** is statistical power to detect true associations. While our Norwegian cohort is, to the best of our knowledge, the largest cohort used for genetic association studies with patients diagnosed by the CCC, it is still substantially underpowered to detect small effect sizes. That is why we added two replication cohorts and meta-analyzed them to increase our overall power. However, there was, at least, one drawback with each of the two replication cohorts. The Danish cohort, while also being diagnosed according to the CCC, was only genotyped for 24 tag SNPs identified in the Norwegian discovery cohort. The UK Biobank data set, on the other hand, while having > 2000 patients and whole-genome data, had less stringent phenotype with self-reported CFS. In order to minimize bias from this large and self-reported cohort (UK Biobank, n=2105) driving the associations signals, we opted to focus on the meta-analysis of top hits identified by the Norwegian discovery cohort. Considering our, overall, low statically power, false negative, as well as false positive associations, cannot be excluded. Furthermore, the SNPs brought forward in our study have all a risk allele frequency <5%. Nevertheless, our study still represents a first step towards understanding the involvement of immunologically relevant genes in ME/CFS.

Prior to our work, studies done by Schlauch et al., Smith et al. and Steiner et al. have all reported significant findings in immune relevant genes like *PTPN22*, *CTLA4* and the *TRA* locus [56, 55, 60]. However, none of these findings were replicated in our larger, statically more representative study.

ME/CFS research lags decades behind other diseases which accumulated tens of thousands of patients in order to establish robust genetic associations. The GWAS era in autoimmune diseases started almost 15 years ago, with a study of 2000 patients for each of the diseases, identifying loci with significance $P<5x10^{-7}$ (i.e. 9 in CD, 3 in RA, 7 in T1D) [213]. Seven years later, an international collaboration involving 29,880 RA cases and 73,758 controls made it possible to establish 101 risk loci with genome-wide significance [214].

In addition to identifying novel, genomic studies in AIDs have revealed that many genetic risk loci are shared between multiple diseases [139, 151, 154, 205, 215, 216]. GWAS are critical to understanding the pleiotropic effects of genes across different AIDs. Already in the early days of GWAS, studies comparing risk loci revealed overlap between AIDs [215]. Diseases like CD and UC share a substantial number of risk loci but relatively few with the other AIDs [216, 217]. This comes as no surprise considering the similar clinical manifestations of CD and UC [218]. On the other hand, RA and T1D share a substantial amount of their risk loci with several other AIDs [215, 160, 142]. Nevertheless, the presence of a shared risk locus does not imply the same underlying causal variant [140]. GWAS loci that harbour alleles with opposite effects in different AIDs have also been reported [219, 220].

Many identified risk loci in AIDs are involved in process of immune cell differentiation and signalling and antigen presentation [139, 215, 221]. A previous candidate SNP study in ME/CFS has reported associations with the immunologically relevant *PTPN22* and *CTLA4* genes [60]. We could, however, not replicate those findings in **paper II.** Overall, findings from **paper II** did not show extensive overlap between ME/CFS and other established AIDs. Nevertheless, the *IGFBP1* region identified in our study was previously reported to be associated with CD [222].

Further fine-mapping of identified loci is essential in order to unravel underlying disease mechanisms which enables those findings to be used in clinical applications and drug development [140]. AID drug targets like TNF blockers, IL23/IL12 antagonists, and CTLA4 agonists, have been shown to reside in GWAS loci of autoimmune diseases [140, 223]. In addition, the combination of approaches, like flow cytometry with GWAS, shows promises in identifying specific proteins, cells, and pathways that could lead to the development of novel drug targets [224].

Our results from **paper II** have to be interpreted in the context of the ME/CFS research field and ought not to be looked at from the genetic perspective of established autoimmune disorders. Even though we did not identify any convincing associations using the iChip in **paper II,** it represents a first step in the right direction when studying ME/CFS genetic architecture and a valuable dataset to be included in future meta-analyses.

5.1.4 No T cell receptor alpha locus associations

While the role of autoreactive T cells in autoimmunity is clearly established, the role of TCR genetic polymorphisms is still unclear. TCR do not only derive their diversity from polymorphisms in TCR loci but rather from somatic recombination. This is in sharp contrast to the HLA genetic region which is the most polymorphic genetic region of the whole genome.

The main aim of **paper III** was to investigate previous claims by Schlauch et al. [56]. This paper is frequently cited in ME/CFS research as an indicator of the involvement of T-cell receptor SNPs in ME/CFS, and it warranted a replication effort as it was based on a small cohort of 40 patients. Additionally, our observed HLA associations from **paper I** naturally also pointed in the direction of TCR investigation.

In total, we analyzed 409 Norwegian ME/CFS (diagnosed with the CCC) and 2105 CFS (self-reported) patients from the UK Biobank for SNPs in the TCR region but did not observe any significant associations. This is in stark contrast to Schlauch et al., who, while using the CCC, only analyzed 40 patients without stated ethnic background. They report GWAS level significance for two SNPs in TRA locus (rs17255510 and rs11157573). While effect sizes are not stated in the paper, calculations based on their data revealed that the two SNPs had an OR>10. Upon deeper investigation of the data presented by Schlauch et al., it became apparent that the observed allele frequencies for their ME/CFS patients did not correspond to our observed frequencies in either the Norwegian or UK cohort (Table 3). In contrast, the two latter datasets showed comparable allele frequencies. Furthermore, our control frequencies were in the same range as that observed in the 1000 genomes dataset. The small sample size (n=40) and perhaps divergent ethnicities used in the Schlauch et al. study, could possibly explain the observed discrepancies in allele frequencies.

Table 3. Overview of the allele frequencies for rs17255510 and rs11157573 from our study and the study by Schlauch et al.

		Norwegian cohort		**UK biobank**		**Schlauch et al.**		**1000G CEU**
SNPs	**Allele**	Cases	Controls	Cases	Controls	Cases	Controls	General
rs17255510	C	0.220	0.218	0.214	0.234	**0.679**	0.171	0.232
rs11157573	C	0.180	0.184	0.171	0.178	**0.488**	0.158	0.197

Nevertheless, both our and the study by Schlauch et al. lack both statistical power and comprehensive genetic investigation of the TRA locus to make any definite claims about TCR involvement in ME/CFS. In addition, several other loci, not yet investigated, are encoding the TCR. The impact of polymorphisms in TCR on disease and immunity is overall a field of investigations that lags behind in all diseases, as these loci are not well captured by neither GWAS nor iChip arrays [225]. Therefore, research in more established diseases, like narcolepsy and MS, still remains inconclusive [226, 227, 228, 229]. Furthermore, investigations of TCR involvement in ME/CFS needs to be expanded into much larger and more stringently diagnosed cohorts in order to observe possible changes in TCR loci.

5.2 Final remarks

The individual studies elaborated in this thesis were intended to investigate different aspects of the core proposition that ME/CFS is an immune-mediated disease. **Papers I and III** investigated genetic changes in key mediators of human autoimmunity, HLA molecules and one TCR gene. The involvement of the HLA polymorphisms is clearly established in autoimmune disease, however, this cannot be said for genetic variation of the TCR loci [99] [230]. Both HLA and TCR loci show, albeit to vastly different extents, allelic sequence variation. Nevertheless, the full scope of TCR polymorphisms and their functional significance in influencing autoimmune disease is unknown [230]. The role of TCR loci involvement, in much more researched diseases like MS [226, 227] and narcolepsy [228, 229], is still unclear. That is why the findings reported by Schlacuh et al. were replication worthy in **paper III**. Even though our power to detect associations was ten times higher than in Schlauch et al., it is still substantially underpowered to detect small effect sizes which are likely to be expected in a complex disorder like ME/CFS.

The lack of reproducible and robust findings is substantial in ME/CFS research. In order to solve this, larger and better-defined cohorts are needed. However, this is not an easy task. Due to the very nature of ME/CFS diagnosis, it is hard to establish well-defined cohorts. This is visible from our work where the Norwegian and Danish cohorts were both diagnosed according to the CCC **(Paper II).** The core symptoms and exclusionary diagnosis are sometimes hard to assess for physicians not routinely dealing with patients with ME/CFS. The assessment of malaise, pain or even psychiatric disorders is to a high degree subjective and prone to error and misdiagnosis. The variability could therefore, even when using the same criteria, be possible significant [10].

Nevertheless**,** projects like DecodeME (www.decodeme.org.uk) will hopefully enable sufficient statistical power to detect genetic changes as well as to enable potential subgroups of ME/CFS patients. An unwanted ally in this undertaking could possibly be the ongoing COVID-19 pandemic. Indeed, according to official USA authorities at the CDC, patients with COVID-19 can develop "a post-viral syndrome that's strikingly similar to myalgic encephalomyelitis/chronic fatigue syndrome". Komaroff et al. have published a paper estimating that the pandemic will generate over 10 million new cases of ME/CFS, globally[231]. In the case of Norway, in particular, the MoBa cohort [232] could be a great resource in following more than 114 000 children, 95 000 mothers and 75 000 fathers and their possible ME/CFS development after COVID-19.

The degree of immune involvement in ME/CFS remains uncertain, however, further studies like ours are warranted to firmly identify the immunological components of ME/CFS. It is important to stress that a sole immune-mediated explanation seems not plausible at this point of research. ME/CFS should rather be investigated in a synergic approach incorporating immunology, genetics, metabolomics and endocrinology. In addition, the impact of ME/CFS on patient's mental health must not be neglected and has to be considered when studying ME/CFS. While psychosomatic or biosocial causes of ME/CFS can be dismissed to a high degree, the relevance of maintaining good mental health for patients cannot be put aside given that the disease takes, sometimes, a severe toll on patients quality of life.

The heterogeneous nature of ME/CFS demands further patient classification and sub-phenotyping in order to establish best patient treatment. Experiences from other complex diseases show that subgrouping of patients is highly relevant for best patient care [233, 234]. Likewise, larger cohorts could reveal that similar clinical manifestations can ultimately have different underlying causes. That is why the establishment of large consortia cohorts is needed. This could, besides other things, reveal important pathophysiological components which lead to the clinical picture we observe in ME/CFS. This need becomes clear when considering that both intravenous cyclophosphamide treatment [184] and cognitive behavioural therapy [235] yield, at least perceived, improvements for some patients currently classified as ME/CFS.

The lack of knowledge in ME/CFS is astonishingly high considering that the disease is studied for over 30 years. Patients suffering is substantial in ME/CFS. The absence of universally accepted diagnostic criteria, as well as any form of standardized care, is something

that encourage more thorough research to narrow down the gap in knowledge. This could, in turn, bring ease to the suffering of millions of patients struggling daily with the effects of ME/CFS.

6 Conclusion

This work presents findings that will hopefully add to the better understanding of ME/CFS. Our primary aim was to investigate the immunogenetic make-up of ME/CFS and to see possible genetic overlap with other autoimmune disorders by conducting a thorough genetic screen of immunological genes in a large ME/CFS patient group diagnosed according to the Canadian consensus criteria.

We showed independent HLA class I and II associations in ME/CFS. However, the iChip analysis showed varying results and emphasized the difficulty in comparing findings between different populations even if diagnosed with the same criteria. In addition, T cell receptor alpha locus involvement in ME/CFS was not replicated in our study and this possibility still remains unclear.

To the best of our knowledge, our study represents the largest and most homogenous genetic study performed in ME/CFS so far. We investigated several immunological relevant genetic regions of the human genome in order to better understand immune involvement in ME/CFS. No definitive statements can be extrapolated from our work, however, the findings we observe indicate, at least in part, the involvement of the immune system in ME/CFS.

References

1. Carruthers, B.M., et al., *Myalgic Encephalomyelitis/Chronic Fatigue Syndrome.* Journal of Chronic Fatigue Syndrome, 2003. **11**(1): p. 7-115.
2. Straus, S.E., *History of Chronic Fatigue Syndrome.* Reviews of Infectious Diseases, 1991. **13**: p. S2-S7.
3. Hashimoto, N., *[History of chronic fatigue syndrome]*. Nihon Rinsho, 2007. **65**(6): p. 975-82.
4. Lim, E.-J. and C.-G. Son, *Review of case definitions for myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS).* Journal of translational medicine, 2020. **18**(1): p. 289-289.
5. Nacul, L.C., et al., *Prevalence of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) in three regions of England: a repeated cross-sectional study in primary care.* BMC medicine, 2011. **9**: p. 91-91.
6. Friedman, K.J., et al., *Our Evolving Understanding of ME/CFS.* Medicina (Kaunas, Lithuania), 2021. **57**(3): p. 200.
7. Son, C.-G., *Differential diagnosis between "chronic fatigue" and "chronic fatigue syndrome".* Integrative medicine research, 2019. **8**(2): p. 89-91.
8. Komaroff, A.L., *Advances in Understanding the Pathophysiology of Chronic Fatigue Syndrome.* Jama, 2019. **322**(6): p. 499-500.
9. Holmes, G.P., et al., *Chronic Fatigue Syndrome: A Working Case Definition.* Annals of Internal Medicine, 1988. **108**(3): p. 387-389.
10. Brurberg, K.G., et al., *Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review.* BMJ Open, 2014. **4**(2): p. e003973.
11. Fukuda, K., et al., *The Chronic Fatigue Syndrome: A Comprehensive Approach to Its Definition and Study.* Annals of Internal Medicine, 1994. **121**(12): p. 953-959.
12. Sharpe, M.C., et al., *A report--chronic fatigue syndrome: guidelines for research.* Journal of the Royal Society of Medicine, 1991. **84**(2): p. 118-121.
13. Brown, A.A., et al., *Contrasting Case Definitions: The ME International Consensus Criteria vs. the Fukuda et al. CFS Criteria.* North American journal of psychology, 2013. **15**(1): p. 103-120.
14. Leonard, A.J., et al., *The Development of a Revised Canadian Myalgic Encephalomyelitis Chronic Fatigue Syndrome Case Definition.* American Journal of Biochemistry and Biotechnology, 2010. **6**(2).
15. Baraniuk, J.N., *Chronic Fatigue Syndrome prevalence is grossly overestimated using Oxford criteria compared to Centers for Disease Control (Fukuda) criteria in a U.S. population study.* Fatigue : biomedicine, health & behavior, 2017. **5**(4): p. 215-230.
16. Lawrie, S.M. and A.J. Pelosi, *Chronic Fatigue Syndrome in the Community Prevalence and Associations.* British Journal of Psychiatry, 1995. **166**(6): p. 793-797.
17. National institute for, h., et al., *Myalgic encephalomyelitis (or encephalopathy)/chronicfatigue syndrome: diagnosis and management.* National institute for health and care excellence, 2021. **11**(6): p. 1115-1126.
18. Jason, L.A., et al., *Case definitions integrating empiric and consensus perspectives.* Fatigue : biomedicine, health & behavior, 2016. **4**(1): p. 1-23.
19. Karfakis, N., *The biopolitics of CFS/ME.* Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 2018. **70**: p. 20-28.
20. Harvey, S.B. and S. Wessely, *Chronic fatigue syndrome: identifying zebras amongst the horses.* BMC medicine, 2009. **7**: p. 58-58.
21. Deary, V., T. Chalder, and M. Sharpe, *The cognitive behavioural model of medically unexplained symptoms: a theoretical and empirical review.* Clin Psychol Rev, 2007. **27**(7): p. 781-97.
22. Jones, J.F., et al., *An evaluation of exclusionary medical/psychiatric conditions in the definition of chronic fatigue syndrome.* BMC Medicine, 2009. **7**(1): p. 57.
23. Jones, J.F., *An extended concept of altered self: Chronic fatigue and post-infection syndromes.* Psychoneuroendocrinology, 2008. **33**(2): p. 119-129.
24. Sotzny, F., et al., *Myalgic Encephalomyelitis/Chronic Fatigue Syndrome – Evidence for an autoimmune disease.* Autoimmunity Reviews, 2018. **17**(6): p. 601-609.

25. Filler, K., et al., *Association of Mitochondrial Dysfunction and Fatigue: A Review of the Literature.* BBA clinical, 2014. **1**: p. 12-23.
26. Brenu, E.W., et al., *Natural killer cells in patients with severe chronic fatigue syndrome.* Auto- immunity highlights, 2013. **4**(3): p. 69-80.
27. Fluge, Ø., et al., *Metabolic profiling indicates impaired pyruvate dehydrogenase function in myalgic encephalopathy/chronic fatigue syndrome.* JCI insight, 2016. **1**(21): p. e89376-e89376.
28. Klimas, N.G., et al., *Immunologic abnormalities in chronic fatigue syndrome.* Journal of clinical microbiology, 1990. **28**(6): p. 1403-1410.
29. Ascherio, A. and K.L. Munger, *EBV and Autoimmunity*, in *Epstein Barr Virus Volume 1: One Herpes Virus: Many Diseases*, C. Münz, Editor. 2015, Springer International Publishing: Cham. p. 365-385.
30. Rasa, S., et al., *Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS).* Journal of translational medicine, 2018. **16**(1): p. 268-268.
31. Schleinitz, N., et al., *Natural killer cells in human autoimmune diseases.* Immunology, 2010. **131**(4): p. 451-458.
32. Kucuksezer, U.C., et al., *The Role of Natural Killer Cells in Autoimmune Diseases.* Frontiers in Immunology, 2021. **12**(79).
33. Eaton-Fitch, N., et al., *A systematic review of natural killer cells profile and cytotoxic function in myalgic encephalomyelitis/chronic fatigue syndrome.* Systematic Reviews, 2019. **8**(1): p. 279.
34. Konstantinov, K., et al., *Autoantibodies to nuclear envelope antigens in chronic fatigue syndrome.* The Journal of clinical investigation, 1996. **98**(8): p. 1888-1896.
35. von Mikecz, A., et al., *High frequency of autoantibodies to insoluble cellular antigens in patients with chronic fatigue syndrome.* Arthritis Rheum, 1997. **40**(2): p. 295-305.
36. Nishikai, M., et al., *Autoantibodies to a 68/48 kDa protein in chronic fatigue syndrome and primary fibromyalgia: a possible marker for hypersomnia and cognitive disorders.* Rheumatology (Oxford), 2001. **40**(7): p. 806-10.
37. Tanaka, S., et al., *Autoantibodies against muscarinic cholinergic receptor in chronic fatigue syndrome.* Int J Mol Med, 2003. **12**(2): p. 225-30.
38. Loebel, M., et al., *Antibodies to β adrenergic and muscarinic cholinergic receptors in patients with Chronic Fatigue Syndrome.* Brain Behav Immun, 2016. **52**: p. 32-39.
39. Moudgil, K.D. and D. Choubey, *Cytokines in autoimmunity: role in induction, regulation, and treatment.* Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research, 2011. **31**(10): p. 695-703.
40. Montoya, J.G., et al., *Cytokine signature associated with disease severity in chronic fatigue syndrome patients.* Proc Natl Acad Sci U S A, 2017. **114**(34): p. E7150-e7158.
41. Roerink, M.E., et al., *Cytokine signatures in chronic fatigue syndrome patients: a Case Control Study and the effect of anakinra treatment.* J Transl Med, 2017. **15**(1): p. 267.
42. Hornig, M., et al., *Distinct plasma immune signatures in ME/CFS are present early in the course of illness.* Science Advances, 2015. **1**(1): p. e1400121.
43. DeBerardinis, R.J. and C.B. Thompson, *Cellular metabolism and disease: what do metabolic outliers teach us?* Cell, 2012. **148**(6): p. 1132-1144.
44. Hoel, F., et al., *A map of metabolic phenotypes in patients with myalgic encephalomyelitis/chronic fatigue syndrome.* JCI Insight, 2021. **6**(16).
45. Yamagiwa, S., et al., *Autoantibodies in primary biliary cirrhosis: recent progress in research on the pathogenetic and clinical significance.* World J Gastroenterol, 2014. **20**(10): p. 2606-12.
46. Lim, E.-J., et al., *Systematic review and meta-analysis of the prevalence of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME).* Journal of Translational Medicine, 2020. **18**(1): p. 100.
47. van Vollenhoven, R.F., *Sex differences in rheumatoid arthritis: more than meets the eye.* BMC Medicine, 2009. **7**(1): p. 12.

48. Bakken, I.J., et al., *Two age peaks in the incidence of chronic fatigue syndrome/myalgic encephalomyelitis: a population-based registry study from Norway 2008-2012.* BMC medicine, 2014. **12**: p. 167-167.
49. Angum, F., et al., *The Prevalence of Autoimmune Disorders in Women: A Narrative Review.* Cureus, 2020. **12**(5): p. e8094-e8094.
50. Albright, F., et al., *Evidence for a heritable predisposition to Chronic Fatigue Syndrome.* BMC neurology, 2011. **11**: p. 62-62.
51. Crawley, E. and G. Davey Smith, *Is chronic fatigue syndrome (CFS/ME) heritable in children, and if so, why does it matter?* Archives of disease in childhood, 2007. **92**(12): p. 1058-1061.
52. Hickie, I.B., et al., *A Twin Study of the Etiology of Prolonged Fatigue and Immune Activation.* Twin Research, 2001. **4**(2): p. 94-102.
53. Schur, E., et al., *Twin Analyses of Fatigue.* Twin Research and Human Genetics, 2007. **10**(5): p. 729-733.
54. Sullivan, P.F., et al., *Twin analyses of chronic fatigue in a Swedish national sample.* Psychological Medicine, 2005. **35**(9): p. 1327-1336.
55. Smith, A.K., et al., *Convergent genomic studies identify association of GRIK2 and NPAS2 with chronic fatigue syndrome.* Neuropsychobiology, 2011. **64**(4): p. 183-94.
56. Schlauch, K.A., et al., *Genome-wide association analysis identifies genetic variations in subjects with myalgic encephalomyelitis/chronic fatigue syndrome.* Translational psychiatry, 2016. **6**(2): p. e730-e730.
57. Perez, M., et al., *Genetic Predisposition for Immune System, Hormone, and Metabolic Dysfunction in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: A Pilot Study.* Frontiers in Pediatrics, 2019. **7**(206).
58. Glusman, G., et al., *Kaviar: an accessible system for testing SNV novelty.* Bioinformatics (Oxford, England), 2011. **27**(22): p. 3216-3217.
59. Goertzel, B.N., et al., *Combinations of single nucleotide polymorphisms in neuroendocrine effector and receptor genes predict chronic fatigue syndrome.* Pharmacogenomics, 2006. **7**(3): p. 475-483.
60. Steiner, S., et al., *Autoimmunity-Related Risk Variants in PTPN22 and CTLA4 Are Associated With ME/CFS With Infectious Onset.* Frontiers in Immunology, 2020. **11**(578).
61. Keller, R.H., et al., *Association Between HLA Class II Antigens and the Chronic Fatigue Immune Dysfunction Syndrome.* Clinical Infectious Diseases, 1994. **18**(Supplement_1): p. S154-S156.
62. Smith, J., et al., *Association of chronic fatigue syndrome with human leucocyte antigen class II alleles.* J Clin Pathol, 2005. **58**(8): p. 860-3.
63. Lande, A., et al., *Human Leukocyte Antigen alleles associated with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS).* Scientific Reports, 2020. **10**(1): p. 5267.
64. Hajdarevic, R., et al., *Fine mapping of the major histocompatibility complex (MHC) in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) suggests involvement of both HLA class I and class II loci.* Brain, Behavior, and Immunity, 2021.
65. Carlo-Stella, N., et al., *Molecular Study of Receptor for Advanced Glycation Endproduct Gene Promoter and Identification of Specific HLA Haplotypes Possibly Involved in Chronic Fatigue Syndrome.* International Journal of Immunopathology and Pharmacology, 2009. **22**(3): p. 745-754.
66. Spitzer, A.R. and M. Broadman, *A Retrospective Review of the Sleep Characteristics in Patients with Chronic Fatigue Syndrome and Fibromyalgia.* Pain Practice, 2010. **10**(4): p. 294-300.
67. Schwartz, R.H., *Historical overview of immunological tolerance.* Cold Spring Harbor perspectives in biology, 2012. **4**(4): p. a006908-a006908.
68. Klein, L., et al., *Antigen presentation in the thymus for positive selection and central tolerance induction.* Nature Reviews Immunology, 2009. **9**(12): p. 833-844.
69. Pelanda, R. and R.M. Torres, *Central B-cell tolerance: where selection begins.* Cold Spring Harbor perspectives in biology, 2012. **4**(4): p. a007146-a007146.

70. Luning Prak, E.T., M. Monestier, and R.A. Eisenberg, *B cell receptor editing in tolerance and autoimmunity.* Annals of the New York Academy of Sciences, 2011. **1217**: p. 96-121.
71. Mueller, D.L., *Mechanisms maintaining peripheral tolerance.* Nature Immunology, 2010. **11**(1): p. 21-27.
72. Wang, L., F.-S. Wang, and M.E. Gershwin, *Human autoimmune diseases: a comprehensive update.* Journal of Internal Medicine, 2015. **278**(4): p. 369-395.
73. Xiao, Z.X., J.S. Miller, and S.G. Zheng, *An updated advance of autoantibodies in autoimmune diseases.* Autoimmunity Reviews, 2021. **20**(2): p. 102743.
74. Suurmond, J. and B. Diamond, *Autoantibodies in systemic autoimmune diseases: specificity and pathogenicity.* The Journal of clinical investigation, 2015. **125**(6): p. 2194-2202.
75. Nicholson, L.B., *The immune system.* Essays in biochemistry, 2016. **60**(3): p. 275-301.
76. Avrameas, S., H. Alexopoulos, and H.M. Moutsopoulos, *Natural Autoantibodies: An Undersugn Hero of the Immune System and Autoimmune Disorders-A Point of View.* Frontiers in immunology, 2018. **9**: p. 1320-1320.
77. Theofilopoulos, A.N., D.H. Kono, and R. Baccala, *The multiple pathways to autoimmunity.* Nature immunology, 2017. **18**(7): p. 716-724.
78. Skapenko, A., et al., *The role of the T cell in autoimmune inflammation.* Arthritis Research & Therapy, 2005. **7**(2): p. S4.
79. Cooper, G.S. and B.C. Stroehla, *The epidemiology of autoimmune diseases.* Autoimmunity Reviews, 2003. **2**(3): p. 119-125.
80. Wirth, K. and C. Scheibenbogen, *A Unifying Hypothesis of the Pathophysiology of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS): Recognitions from the finding of autoantibodies against ß2-adrenergic receptors.* Autoimmun Rev, 2020. **19**(6): p. 102527.
81. Mandarano, A.H., et al., *Myalgic encephalomyelitis/chronic fatigue syndrome patients exhibit altered T cell metabolism and cytokine associations.* J Clin Invest, 2020. **130**(3): p. 1491-1505.
82. Rivas, J.L., et al., *Association of T and NK Cell Phenotype With the Diagnosis of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS).* Front Immunol, 2018. **9**: p. 1028.
83. Smith, D.A. and D.R. Germolec, *Introduction to immunology and autoimmunity.* Environmental health perspectives, 1999. **107 Suppl 5**(Suppl 5): p. 661-665.
84. Lee, K.H., et al., *Understanding the immunopathogenesis of autoimmune diseases by animal studies using gene modulation: A comprehensive review.* Autoimmun Rev, 2020. **19**(3): p. 102469.
85. Lafaille, J.J., *The role of helper T cell subsets in autoimmune diseases.* Cytokine Growth Factor Rev, 1998. **9**(2): p. 139-51.
86. Gensous, N., et al., *T Follicular Helper Cells in Autoimmune Disorders.* Frontiers in Immunology, 2018. **9**(1637).
87. Rao, D.A., *The rise of peripheral T helper cells in autoimmune disease.* Nature Reviews Rheumatology, 2019. **15**(8): p. 453-454.
88. Frommer, L. and G.J. Kahaly, *Type 1 Diabetes and Autoimmune Thyroid Disease-The Genetic Link.* Frontiers in endocrinology, 2021. **12**: p. 618213-618213.
89. Bradfield, J.P., et al., *A Genome-Wide Meta-Analysis of Six Type 1 Diabetes Cohorts Identifies Multiple Associated Loci.* PLOS Genetics, 2011. **7**(9): p. e1002293.
90. DiMeglio, L.A., C. Evans-Molina, and R.A. Oram, *Type 1 diabetes.* Lancet (London, England), 2018. **391**(10138): p. 2449-2462.
91. Cooper, G.S., M.L.K. Bynum, and E.C. Somers, *Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases.* Journal of autoimmunity, 2009. **33**(3-4): p. 197-207.
92. Ellrichmann, M., et al., *Endoscopic ultrasound of the colon for the differentiation of Crohn's disease and ulcerative colitis in comparison with healthy controls.* Aliment Pharmacol Ther, 2014. **39**(8): p. 823-33.
93. Kumar, B.V., T.J. Connors, and D.L. Farber, *Human T Cell Development, Localization, and Function throughout Life.* Immunity, 2018. **48**(2): p. 202-213.
94. Pilli, D., et al., *Expanding Role of T Cells in Human Autoimmune Diseases of the Central Nervous System.* Frontiers in Immunology, 2017. **8**(652).

95. Dornmair, K., et al., *T-cell-mediated autoimmunity: novel techniques to characterize autoreactive T-cell receptors.* The American journal of pathology, 2003. **163**(4): p. 1215-1226.
96. Kelly, A. and J. Trowsdale, *Introduction: MHC/KIR and governance of specificity.* Immunogenetics, 2017. **69**(8-9): p. 481-488.
97. Szeto, C., et al., *TCR Recognition of Peptide-MHC-I: Rule Makers and Breakers.* International journal of molecular sciences, 2020. **22**(1): p. 68.
98. Lefranc, M.-P., *Immunoglobulin and T Cell Receptor Genes: IMGT(®) and the Birth and Rise of Immunoinformatics.* Frontiers in immunology, 2014. **5**: p. 22-22.
99. Trowsdale, J. and J.C. Knight, *Major histocompatibility complex genomics and human disease.* Annual review of genomics and human genetics, 2013. **14**: p. 301-323.
100. Jung, D. and F.W. Alt, *Unraveling V(D)J Recombination: Insights into Gene Regulation.* Cell, 2004. **116**(2): p. 299-311.
101. Bluestone, J.A., et al., *T cells in the control of organ-specific autoimmunity.* The Journal of clinical investigation, 2015. **125**(6): p. 2250-2260.
102. Raphael, I., R.R. Joern, and T.G. Forsthuber, *Memory CD4(+) T Cells in Immunity and Autoimmune Diseases.* Cells, 2020. **9**(3): p. 531.
103. Parham, P. and T. Ohta, *Population Biology of Antigen Presentation by MHC Class I Molecules.* Science, 1996. **272**(5258): p. 67.
104. Wosen, J.E., et al., *Epithelial MHC Class II Expression and Its Role in Antigen Presentation in the Gastrointestinal and Respiratory Tracts.* Frontiers in Immunology, 2018. **9**(2144).
105. Robinson, J., et al., *IPD-IMGT/HLA Database.* Nucleic Acids Res, 2020. **48**(D1): p. D948-d955.
106. Dendrou, C.A., et al., *HLA variation and disease.* Nature Reviews Immunology, 2018. **18**: p. 325.
107. Brewerton, D.A., et al., *ACUTE ANTERIOR UVEITIS AND HL-A 27.* The Lancet, 1973. **302**(7836): p. 994-996.
108. Falchuk, Z.M., G.N. Rogentine, and W. Strober, *Predominance of histocompatibility antigen HL-A8 in patients with gluten-sensitive enteropathy.* The Journal of clinical investigation, 1972. **51**(6): p. 1602-1605.
109. Matzaraki, V., et al., *The MHC locus and genetic susceptibility to autoimmune and infectious diseases.* Genome biology, 2017. **18**(1): p. 76-76.
110. Phloyphisut, P., et al., *MHCSeqNet: a deep neural network model for universal MHC binding prediction.* BMC Bioinformatics, 2019. **20**(1): p. 270.
111. Motyer, A., et al., *Practical Use of Methods for Imputation of HLA Alleles from SNP Genotype Data.* bioRxiv, 2016: p. 091009.
112. Kurkó, J., et al., *Genetics of rheumatoid arthritis - a comprehensive review.* Clinical reviews in allergy & immunology, 2013. **45**(2): p. 170-179.
113. Zhernakova, A., et al., *Meta-Analysis of Genome-Wide Association Studies in Celiac Disease and Rheumatoid Arthritis Identifies Fourteen Non-HLA Shared Loci.* PLOS Genetics, 2011. **7**(2): p. e1002004.
114. Julià, A., et al., *Genome-wide association study meta-analysis identifies five new loci for systemic lupus erythematosus.* Arthritis Research & Therapy, 2018. **20**(1): p. 100.
115. Tsoi, L.C., et al., *Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants.* Nature communications, 2017. **8**: p. 15382-15382.
116. Tizaoui, K., *Multiple sclerosis genetics: Results from meta-analyses of candidate-gene association studies.* Cytokine, 2018. **106**: p. 154-164.
117. Arase, H., *MHC: new insights from old molecules into the pathogenesis of autoimmunity.* International Immunology, 2021.
118. Salzberg, S.L., *Open questions: How many genes do we have?* BMC biology, 2018. **16**(1): p. 94-94.
119. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
120. Eichler, E.E., *Genetic Variation, Comparative Genomics, and the Diagnosis of Disease.* The New England journal of medicine, 2019. **381**(1): p. 64-74.

121. Crawford, D.C. and D.A. Nickerson, *Definition and clinical importance of haplotypes.* Annu Rev Med, 2005. **56**: p. 303-20.
122. Hamazaki, K. and H. Iwata, *RAINBOW: Haplotype-based genome-wide association study using a novel SNP-set method.* PLOS Computational Biology, 2020. **16**(2): p. e1007663.
123. Spataro, N., et al., *Properties of human disease genes and the role of genes linked to Mendelian disorders in complex disease aetiology.* Human molecular genetics, 2017. **26**(3): p. 489-500.
124. Motulsky, A.G., *Genetics of complex diseases.* Journal of Zhejiang University. Science. B, 2006. **7**(2): p. 167-168.
125. Sollid, L.M. and B. Jabri, *Triggers and drivers of autoimmunity: lessons from coeliac disease.* Nature reviews. Immunology, 2013. **13**(4): p. 294-302.
126. Lewis, D., J. Haridy, and E.D. Newnham, *Testing for coeliac disease.* Australian prescriber, 2017. **40**(3): p. 105-108.
127. Sams, A. and J. Hawks, *Celiac disease as a model for the evolution of multifactorial disease in humans.* Hum Biol, 2014. **86**(1): p. 19-36.
128. Slatkin, M., *Linkage disequilibrium--understanding the evolutionary past and mapping the medical future.* Nature reviews. Genetics, 2008. **9**(6): p. 477-485.
129. Ardlie, K.G., L. Kruglyak, and M. Seielstad, *Patterns of linkage disequilibrium in the human genome.* Nature Reviews Genetics, 2002. **3**(4): p. 299-309.
130. VanLiere, J.M. and N.A. Rosenberg, *Mathematical properties of the r2 measure of linkage disequilibrium.* Theoretical population biology, 2008. **74**(1): p. 130-137.
131. Blomhoff, A., et al., *Linkage disequilibrium and haplotype blocks in the MHC vary in an HLA haplotype specific manner assessed mainly by DRB1*03 and DRB1*04 haplotypes.* Genes & Immunity, 2006. **7**(2): p. 130-140.
132. Chang, M., L. He, and L. Cai, *An Overview of Genome-Wide Association Studies.* Methods Mol Biol, 2018. **1754**: p. 97-108.
133. Visscher, P.M., et al., *10 Years of GWAS Discovery: Biology, Function, and Translation.* Am J Hum Genet, 2017. **101**(1): p. 5-22.
134. Korte, A. and A. Farlow, *The advantages and limitations of trait analysis with GWAS: a review.* Plant methods, 2013. **9**: p. 29-29.
135. Cano-Gamez, E. and G. Trynka, *From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases.* Frontiers in Genetics, 2020. **11**(424).
136. Jordan, B., *Genes and non-mendelian diseases: dealing with complexity.* Perspect Biol Med, 2014. **57**(1): p. 118-31.
137. Boyle, E.A., Y.I. Li, and J.K. Pritchard, *An Expanded View of Complex Traits: From Polygenic to Omnigenic.* Cell, 2017. **169**(7): p. 1177-1186.
138. Maurano, M.T., et al., *Systematic localization of common disease-associated variation in regulatory DNA.* Science (New York, N.Y.), 2012. **337**(6099): p. 1190-1195.
139. Ye, J., K.M. Gillespie, and S. Rodriguez, *Unravelling the Roles of Susceptibility Loci for Autoimmune Diseases in the Post-GWAS Era.* Genes, 2018. **9**(8): p. 377.
140. Caliskan, M., C.D. Brown, and J.C. Maranville, *A catalog of GWAS fine-mapping efforts in autoimmune disease.* The American Journal of Human Genetics, 2021. **108**(4): p. 549-563.
141. Kochi, Y., *Genetics of autoimmune diseases: perspectives from genome-wide association studies.* International immunology, 2016. **28**(4): p. 155-161.
142. Okada, Y., et al., *Genetics of rheumatoid arthritis: 2018 status.* Annals of the Rheumatic Diseases, 2019. **78**(4): p. 446.
143. Cerqueira, J.X.M., et al., *Independent and cumulative coeliac disease-susceptibility loci are associated with distinct disease phenotypes.* Journal of Human Genetics, 2021. **66**(6): p. 613-623.
144. Qu, H.-Q., et al., *Genetic architecture of type 1 diabetes with low genetic risk score informed by 41 unreported loci.* Communications Biology, 2021. **4**(1): p. 908.
145. Cárdenas-Roldán, J., A. Rojas-Villarraga, and J.-M. Anaya, *How do autoimmune diseases cluster in families? A systematic review and meta-analysis.* BMC medicine, 2013. **11**: p. 73-73.

146. Anaya, J.M., Y. Shoenfeld, and R. Cervera, *Facts and challenges for the autoimmunologist. Lessons from the second Colombian autoimmune symposium.* Autoimmun Rev, 2012. **11**(4): p. 249-51.
147. Mills, M.C. and C. Rahal, *A scientometric review of genome-wide association studies.* Communications biology, 2019. **2**: p. 9-9.
148. Alshobaili, H.A., et al., *Genetic background of psoriasis.* International journal of health sciences, 2010. **4**(1): p. 23-29.
149. Reveille, J.D., *Major histocompatibility genes and ankylosing spondylitis.* Best Pract Res Clin Rheumatol, 2006. **20**(3): p. 601-9.
150. International Genetics of Ankylosing Spondylitis, C., et al., *Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci.* Nat Genet, 2013. **45**(7): p. 730-8.
151. Serrano, N.C., P. Millan, and M.C. Páez, *Non-HLA associations with autoimmune diseases.* Autoimmun Rev, 2006. **5**(3): p. 209-14.
152. Li, Y.R., et al., *Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases.* Nature Medicine, 2015. **21**(9): p. 1018-1027.
153. Ramos, P.S., A.M. Shedlock, and C.D. Langefeld, *Genetics of autoimmune diseases: insights from population genetics.* Journal of Human Genetics, 2015. **60**(11): p. 657-664.
154. Kreiner, E., et al., *Shared genetic variants suggest common pathways in allergy and autoimmune diseases.* J Allergy Clin Immunol, 2017. **140**(3): p. 771-781.
155. Richard-Miceli, C. and L.A. Criswell, *Emerging patterns of genetic overlap across autoimmune disorders.* Genome Medicine, 2012. **4**(1): p. 6.
156. Smyth, D.J., et al., *Shared and distinct genetic variants in type 1 diabetes and celiac disease.* N Engl J Med, 2008. **359**(26): p. 2767-77.
157. Coenen, M.J., et al., *Common and different genetic background for rheumatoid arthritis and coeliac disease.* Hum Mol Genet, 2009. **18**(21): p. 4195-203.
158. Festen, E.A., et al., *A meta-analysis of genome-wide association scans identifies IL18RAP, PTPN2, TAGAP, and PUS10 as shared risk loci for Crohn's disease and celiac disease.* PLoS Genet, 2011. **7**(1): p. e1001283.
159. Ramos, P.S., et al., *A Comprehensive Analysis of Shared Loci between Systemic Lupus Erythematosus (SLE) and Sixteen Autoimmune Diseases Reveals Limited Genetic Overlap.* PLOS Genetics, 2011. **7**(12): p. e1002406.
160. Márquez, A. and J. Martín, *Genetic overlap between type 1 diabetes and other autoimmune diseases.* Seminars in Immunopathology, 2021.
161. Manchia, M., et al., *The Impact of Phenotypic and Genetic Heterogeneity on Results of Genome Wide Association Studies of Complex Diseases.* PLOS ONE, 2013. **8**(10): p. e76295.
162. Cullinan, J., et al., *Perceptions of European ME/CFS Experts Concerning Knowledge and Understanding of ME/CFS among Primary Care Physicians in Europe: A Report from the European ME/CFS Research Network (EUROMENE).* Medicina (Kaunas), 2021. **57**(3).
163. Teo, Y.Y., *Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure.* Curr Opin Lipidol, 2008. **19**(2): p. 133-43.
164. Gaspar, H.A. and G. Breen, *Probabilistic ancestry maps: a method to assess and visualize population substructures in genetics.* BMC Bioinformatics, 2019. **20**(1): p. 116.
165. Beecham, A.H., et al., *Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis.* Nature Genetics, 2013. **45**(11): p. 1353-1360.
166. Liu, J.Z., et al., *Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis.* Nature Genetics, 2013. **45**(6): p. 670-675.
167. Barrett, J.C., et al., *Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes.* Nat Genet, 2009. **41**(6): p. 703-7.
168. Cortes, A. and M.A. Brown, *Promise and pitfalls of the Immunochip.* Arthritis Res Ther, 2011. **13**(1): p. 101.
169. Verlouw, J.A.M., et al., *A comparison of genotyping arrays.* European Journal of Human Genetics, 2021.

170. Shi, S., et al., *Comprehensive Assessment of Genotype Imputation Performance.* Human Heredity, 2018. **83**(3): p. 107-116.
171. Das, S., et al., *Next-generation genotype imputation service and methods.* Nature genetics, 2016. **48**(10): p. 1284-1287.
172. Zheng, H.-F., et al., *Effect of Genome-Wide Genotyping and Reference Panels on Rare Variants Imputation.* Journal of Genetics and Genomics, 2012. **39**(10): p. 545-550.
173. Zhang, B., et al., *Practical Consideration of Genotype Imputation: Sample Size, Window Size, Reference Choice, and Untyped Rate.* Stat Interface, 2011. **4**(3): p. 339-352.
174. Anderson, C.A., et al., *Data quality control in genetic case-control association studies.* Nature Protocols, 2010. **5**(9): p. 1564-1573.
175. Bravo-Egana, V., H. Sanders, and N. Chitnis, *New challenges, new opportunities: Next generation sequencing and its place in the advancement of HLA typing.* Human Immunology, 2021. **82**(7): p. 478-487.
176. Dunckley, H., *HLA typing by SSO and SSP methods.* Methods Mol Biol, 2012. **882**: p. 9-25.
177. Erlich, R.L., et al., *Next-generation sequencing for HLA typing of class I loci.* BMC Genomics, 2011. **12**(1): p. 42.
178. Creary, L.E., et al., *Next-Generation Sequencing Identifies Extended HLA Class I and II Haplotypes Associated With Early-Onset and Late-Onset Myasthenia Gravis in Italian, Norwegian, and Swedish Populations.* Frontiers in Immunology, 2021. **12**(1754).
179. Zeggini, E. and J.P.A. Ioannidis, *Meta-analysis in genome-wide association studies.* Pharmacogenomics, 2009. **10**(2): p. 191-201.
180. Thompson, J.R., J. Attia, and C. Minelli, *The meta-analysis of genome-wide association studies.* Briefings in Bioinformatics, 2011. **12**(3): p. 259-269.
181. Panagiotou, O.A., et al., *The Power of Meta-Analysis in Genome-Wide Association Studies.* Annual Review of Genomics and Human Genetics, 2013. **14**(1): p. 441-465.
182. Kaler, A.S. and L.C. Purcell, *Estimation of a significance threshold for genome-wide association studies.* BMC Genomics, 2019. **20**(1): p. 618.
183. Pe'er, I., et al., *Estimation of the multiple testing burden for genomewide association studies of nearly all common variants.* Genet Epidemiol, 2008. **32**(4): p. 381-5.
184. Rekeland, I.G., et al., *Intravenous Cyclophosphamide in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. An Open-Label Phase II Study.* Frontiers in medicine, 2020. **7**: p. 162-162.
185. Ahmed, A.R. and S.M. Hombal, *Cyclophosphamide (Cytoxan): A review on relevant pharmacology and clinical uses.* Journal of the American Academy of Dermatology, 1984. **11**(6): p. 1115-1126.
186. Sciurti, M., et al., *Genetic susceptibilty and celiac disease: what role do HLA haplotypes play?* Acta bio-medica : Atenei Parmensis, 2018. **89**(9-S): p. 17-21.
187. Aletaha, D., et al., *2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative.* Ann Rheum Dis, 2010. **69**(9): p. 1580-8.
188. Fanouriakis, A., et al., *2019 update of the EULAR recommendations for the management of systemic lupus erythematosus.* Annals of the Rheumatic Diseases, 2019. **78**(6): p. 736.
189. Al-Toma, A., et al., *European Society for the Study of Coeliac Disease (ESsCD) guideline for coeliac disease and other gluten-related disorders.* United European Gastroenterology Journal, 2019. **7**(5): p. 583-613.
190. Moretta, L. and A. Moretta, *Killer immunoglobulin-like receptors.* Curr Opin Immunol, 2004. **16**(5): p. 626-33.
191. Agrawal, S. and S. Prakash, *Significance of KIR like natural killer cell receptors in autoimmune disorders.* Clinical Immunology, 2020. **216**: p. 108449.
192. Pasi, A., et al., *Excess of activating killer cell immunoglobulin-like receptors and lack of HLA-Bw4 ligands: A two-edged weapon in chronic fatigue syndrome.* Mol Med Rep, 2011. **4**(3): p. 535-540.
193. Kawabata, Y., et al., *Genome-Wide Association Study Confirming a Strong Effect of HLA and Identifying Variants in CSAD/lnc-ITGB7-1 on Chromosome 12q13.13 Associated With Susceptibility to Fulminant Type 1 Diabetes.* Diabetes, 2019. **68**(3): p. 665.

194. Okada, Y., et al., *HLA-Cw*1202-B*5201-DRB1*1502 Haplotype Increases Risk for Ulcerative Colitis but Reduces Risk for Crohn's Disease.* Gastroenterology, 2011. **141**(3): p. 864-871.e5.
195. Estrada, K., et al., *A whole-genome sequence study identifies genetic risk factors for neuromyelitis optica.* Nature Communications, 2018. **9**(1): p. 1929.
196. Thorsby, E. and B.A. Lie, *HLA associated genetic predisposition to autoimmune diseases: Genes involved and possible mechanisms.* Transplant Immunology, 2005. **14**(3): p. 175-182.
197. Parzanese, I., et al., *Celiac disease: From pathophysiology to treatment.* World journal of gastrointestinal pathophysiology, 2017. **8**(2): p. 27-38.
198. Bodis, G., V. Toth, and A. Schwarting, *Role of Human Leukocyte Antigens (HLA) in Autoimmune Diseases.* Rheumatology and therapy, 2018. **5**(1): p. 5-20.
199. O'Dell, J.R., et al., *HLA-DRB1 typing in rheumatoid arthritis: predicting response to specific treatments.* Ann Rheum Dis, 1998. **57**(4): p. 209-13.
200. Wysocki, T., M. Olesińska, and A. Paradowska-Gorycka, *Current Understanding of an Emerging Role of HLA-DRB1 Gene in Rheumatoid Arthritis-From Research to Clinical Practice.* Cells, 2020. **9**(5): p. 1127.
201. Airaksinen, L., et al., *Influence of HLA-DQ2.5 Dose on Clinical Picture of Unrelated Celiac Disease Patients.* Nutrients, 2020. **12**(12).
202. Vader, W., et al., *The HLA-DQ2 gene dose effect in celiac disease is directly related to the magnitude and breadth of gluten-specific T cell responses.* Proc Natl Acad Sci U S A, 2003. **100**(21): p. 12390-5.
203. Gutiérrez, S., et al., *The human digestive tract has proteases capable of gluten hydrolysis.* Molecular metabolism, 2017. **6**(7): p. 693-702.
204. Márquez, A., et al., *Meta-analysis of Immunochip data of four autoimmune diseases reveals novel single-disease and cross-phenotype associations.* Genome Medicine, 2018. **10**(1): p. 97.
205. Ricaño-Ponce, I., et al., *Immunochip meta-analysis in European and Argentinian populations identifies two novel genetic loci associated with celiac disease.* Eur J Hum Genet, 2020. **28**(3): p. 313-323.
206. Fu, M.M., et al., *The Golgi Outpost Protein TPPP Nucleates Microtubules and Is Critical for Myelination.* Cell, 2019. **179**(1): p. 132-146.e14.
207. Höftberger, R., et al., *Tubulin polymerization promoting protein (TPPP/p25) as a marker for oligodendroglial changes in multiple sclerosis.* Glia, 2010. **58**(15): p. 1847-57.
208. Xu, W., et al., *Dysregulation of Rab5-mediated endocytic pathways in Alzheimer's disease.* Traffic, 2018. **19**(4): p. 253-262.
209. Heneka, M.T., et al., *Neuroinflammation in Alzheimer's disease.* The Lancet. Neurology, 2015. **14**(4): p. 388-405.
210. Nakatomi, Y., H. Kuratsune, and Y. Watanabe, *[Neuroinflammation in the Brain of Patients with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome].* Brain Nerve, 2018. **70**(1): p. 19-25.
211. Nakatomi, Y., et al., *Neuroinflammation in Patients with Chronic Fatigue Syndrome/Myalgic Encephalomyelitis: An ^{11}C-(R)-PK11195 PET Study.* J Nucl Med, 2014. **55**(6): p. 945-50.
212. VanElzakker, M.B., S.A. Brumfield, and P.S. Lara Mejia, *Neuroinflammation and Cytokines in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS): A Critical Review of Research Methods.* Frontiers in Neurology, 2019. **9**(1033).
213. Wellcome Trust Case Control, C., *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-678.
214. Okada, Y., et al., *Genetics of rheumatoid arthritis contributes to biology and drug discovery.* Nature, 2014. **506**(7488): p. 376-81.
215. Zhernakova, A., C.C. van Diemen, and C. Wijmenga, *Detecting shared pathogenesis from the shared genetics of immune-related diseases.* Nature Reviews Genetics, 2009. **10**(1): p. 43-55.
216. Anderson, C.A., et al., *Investigation of Crohn's disease risk loci in ulcerative colitis further defines their molecular relationship.* Gastroenterology, 2009. **136**(2): p. 523-9.e3.
217. Yang, Y., et al., *Investigating the shared genetic architecture between multiple sclerosis and inflammatory bowel diseases.* Nature Communications, 2021. **12**(1): p. 5641.

218. Roushan, N., et al., *Differentiation of Crohn's disease and ulcerative colitis using intestinal wall thickness of the colon: A Diagnostic accuracy study of endoscopic ultrasonography*. Medical journal of the Islamic Republic of Iran, 2019. **33**: p. 57-57.
219. González-Serna, D., et al., *A cross-disease meta-GWAS identifies four new susceptibility loci shared between systemic sclerosis and Crohn's disease.* Sci Rep, 2020. **10**(1): p. 1862.
220. Wang, K., et al., *Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects.* Hum Mol Genet, 2010. **19**(10): p. 2059-67.
221. Ding, J., A. Frantzeskos, and G. Orozco, *Functional interrogation of autoimmune disease genetics using CRISPR/Cas9 technologies and massively parallel reporter assays.* Seminars in Immunopathology, 2021.
222. Ye, B.D., et al., *Association of FUT2 and ABO with Crohn's disease in Koreans.* J Gastroenterol Hepatol, 2020. **35**(1): p. 104-109.
223. Buniello, A., et al., *The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019.* Nucleic Acids Res, 2019. **47**(D1): p. D1005-d1012.
224. Orrù, V., et al., *Application of Genetic Studies to Flow Cytometry Data and Its Impact on Therapeutic Intervention for Autoimmune Disease.* Frontiers in Immunology, 2021. **12**(3513).
225. Mikocziova, I., V. Greiff, and L.M. Sollid, *Immunoglobulin germline gene variation and its impact on human disease.* Genes & Immunity, 2021. **22**(4): p. 205-217.
226. Cotsapas, C. and M. Mitrovic, *Genome-wide association studies of multiple sclerosis.* Clinical & Translational Immunology, 2018. **7**(6): p. e1018.
227. Watson, C.T., et al., *Revisiting the T-cell receptor alpha/delta locus and possible associations with multiple sclerosis.* Genes & Immunity, 2011. **12**(2): p. 59-66.
228. Hallmayer, J., et al., *Narcolepsy is strongly associated with the T-cell receptor alpha locus.* Nature genetics, 2009. **41**(6): p. 708-711.
229. Ouyang, H., et al., *Relationship between T-cell receptor α gene polymorphisms and symptomatic differences in patients with narcolepsy type 1.* Chinese Medical Journal, 2019. **132**(15).
230. Gras, S., et al., *Allelic polymorphism in the T cell receptor and its impact on immune responses.* The Journal of experimental medicine, 2010. **207**(7): p. 1555-1567.
231. Komaroff, A.L. and L. Bateman, *Will COVID-19 Lead to Myalgic Encephalomyelitis/Chronic Fatigue Syndrome?* Frontiers in Medicine, 2021. **7**(1132).
232. Magnus, P., et al., *Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa).* Int J Epidemiol, 2016. **45**(2): p. 382-8.
233. Luurssen-Masurel, N., et al., *The impact of different (rheumatoid) arthritis phenotypes on patients' lives.* Rheumatology, 2021. **60**(8): p. 3716-3726.
234. Cortesi, P.A., et al., *The Prevalence and Treatment Status of Different Multiple Sclerosis Phenotypes in a Italian Reference Center.* Value in Health, 2017. **20**(9): p. A720.
235. Larun, L., et al., *Exercise therapy for chronic fatigue syndrome.* Cochrane Database Syst Rev, 2019. **10**(10): p. Cd003200.

I

Brain, Behavior, and Immunity 98 (2021) 101–109

Contents lists available at ScienceDirect

Brain Behavior and Immunity

journal homepage: www.elsevier.com/locate/ybrbi

Fine mapping of the major histocompatibility complex (MHC) in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) suggests involvement of both HLA class I and class II loci

Riad Hajdarevic [a,b,*], Asgeir Lande [a,b], Ingrid Rekeland [c], Anne Rydland [a,b], Elin B. Strand [d,e], Daisy D. Sosa [d,f], Lisa E Creary [g,h], Olav Mella [c,i], Torstein Egeland [b,j], Ola D. Saugstad [k], Øystein Fluge [c,i], Benedicte A. Lie [a,b,j,1], Marte K. Viken [a,j,*,1]

[a] Department of Medical Genetics, University of Oslo and Oslo University Hospital, Oslo, Norway
[b] Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[c] Department of Oncology and Medical Physics, Haukeland University Hospital, Bergen, Norway
[d] CFS/ME Center, Oslo University Hospital, Oslo, Norway
[e] Faculty of Health, VID Specialized University, Stavanger, Norway
[f] National Advisory Unit on CFS/ME, Oslo, Norway
[g] Department of Pathology, Stanford University, School of Medicine, Palo Alto, CA, USA
[h] Histocompatibility and Immunogenetics Laboratory, Stanford Blood Center, Palo Alto, CA, USA
[i] Department of Clinical Science, University of Bergen, Norway
[j] Department of Immunology, Oslo University Hospital, Oslo, Norway
[k] Department of Pediatric Research, Oslo University Hospital, University of Oslo, Oslo, Norway

ARTICLE INFO

Keywords:
ME/CFS
Human leukocyte antigen
HLA fine-mapping

ABSTRACT

The etiology of myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) is unknown, but involvement of the immune system is one of the proposed underlying mechanisms. Human leukocyte antigen (HLA) associations are hallmarks of immune-mediated and autoimmune diseases. We have previously performed high resolution HLA genotyping and detected associations between ME/CFS and certain HLA class I and class II alleles. However, the HLA complex harbors numerous genes of immunological importance, and there is extensive and complex linkage disequilibrium across the region. In the current study, we aimed to fine map the association signals in the HLA complex by genotyping five additional classical HLA loci and 5,342 SNPs in 427 Norwegian ME/CFS patients, diagnosed according to the Canadian Consensus Criteria, and 480 healthy Norwegian controls. SNP association analysis revealed two distinct and independent association signals ($p \leq 0.001$) tagged by rs4711249 in the HLA class I region and rs9275582 in the HLA class II region. Furthermore, the primary association signal in the HLA class II region was located within the HLA-DQ gene region, most likely due to HLA-DQB1, particularly the amino acid position 57 (aspartic acid/alanine) in the peptide binding groove, or an intergenic SNP upstream of *HLA-DQB1*. In the HLA class I region, the putative causal locus might map outside the classical HLA genes as the association signal spans several genes (*DDR1*, *GTF2H4*, *VARS2*, *SFTA2* and *DPCR1*) with expression levels influenced by the ME/CFS associated SNP genotype. Taken together, our results implicate the involvement of the MHC, and in particular the HLA-DQB1 gene, in ME/CFS. These findings should be replicated in larger cohorts, particularly to verify the putative involvement of *HLA-DQB1*, a gene important for antigen-presentation to T cells and known to harbor alleles providing the largest risk for well–established autoimmune diseases.

1. Introduction

Myalgic encephalomyelitis or chronic fatigue syndrome (ME/CFS) is a disease of unknown etiology and pathogenesis resulting in a variety of symptoms like post exertional malaise (PEM), brain fog, fatigue and pain. No drug interventions are currently available, thus, emphasizing

* Corresponding authors at: Department of Medical Genetics, Oslo University Hospital, PO box 4956 Nydalen, 0424 Oslo, Norway.
E-mail addresses: riad.hajdarevic@medisin.uio.no (R. Hajdarevic), m.k.viken@ous-research.no (M.K. Viken).
[1] Joint last authors.

https://doi.org/10.1016/j.bbi.2021.08.219
Received 23 June 2021; Received in revised form 4 August 2021; Accepted 9 August 2021
Available online 14 August 2021

the need for more research. Comorbidity of ME/CFS with various autoimmune diseases is observed in a substantial number of patients, likewise for fibromyalgia and postural orthostatic tachycardia syndrome (POTS) (Sotzny et al., 2018). The presence of hereditary components in ME/CFS susceptibility is suggested by an increased disease risk in first degree relatives (Lawrie and Pelosi, 1995; Buchwald, 2001; Carruthers et al., 2003; Albright et al., 2011).

An increasing amount of evidence points towards an immune and/or metabolically mediated cause for ME/CFS (Sepúlveda et al., 2019; Smith, 2005; Sotzny et al., 2018). Immune alterations, similar to those of established autoimmune diseases (Sotzny et al., 2018), as well as onset after viral exposure to e.g Epstein-Barr-virus or human papillomavirus, T cell alterations and presence of autoantibodies, have all been suggested in ME/CFS (Sepúlveda et al., 2019; Sotzny et al., 2018; Rasa et al., 2018; Feiring et al., 2017). The major genetic determinants for most autoimmune diseases are the human leukocyte antigen (HLA) genes within the Major Histocompatibility complex (MHC) (Trowsdale and Knight, 2013). This genetic region stretches over 3.7 million base pairs on chromosome 6p21 and consists of>250 genes. The region occupies approximately 0.13% of the human genome but contains ~0.5% (>150) of all protein coding genes (Trowsdale and Knight, 2013). >60% of these genes encode immunologically important proteins (Trowsdale and Knight, 2013). The extended MHC (7.5 Mb) is typically divided into the extended class I, class I, class III, class II and extended class II subregions, with the classical HLA genes encoding for antigen presenting molecules located in the class I and II subregions. These genes harbor variants associated with both susceptibility and protection for numerous autoimmune diseases (Trowsdale and Knight, 2013), however, the hyperpolymorphic properties and extensive linkage disequilibrium (LD) in the region can make it difficult to pinpoint causal genetic variants (Thorsby and Lie, 2005).

In the case of ME/CFS, investigation by our group and others have previously identified associations with specific HLA alleles (Keller, 1994; Smith, 2005; Lande, 2020). The studies that investigated the involvement of MHC in ME/CFS are inconsistent in regard to the patient inclusion criteria and HLA typing methodology (Keller, 1994; Smith, 2005; Underhill et al., 2001; Helbig et al., 2003; Ortega-Hernandez et al., 1173; Pasi, 2011), and reproducible associations are lacking. In the largest study (N = 110), prior to our previous study, the most significant association was with the serologically defined HLA-DQ3 (OR = 1.8, 95% CI 1.2–2.8) (Keller, 1994). Furthermore, associations with HLA alleles *HLA-DQA1*01*, *HLA-DRB1*13:01* and *HLA-DQB1*06:02* have also been reported (Smith, 2005; Carlo-Stella et al., 2009; Spitzer and Broadman, 2010). Most of these studies incorporate<50 patients and are therefore statistically underpowered for detecting associations of moderate effect size.

In our previous study by Lande (2020), we found a significant ME/CFS association with *HLA-DQB1*03:03* and *HLA-C*07:04* alleles. A subsequent study observed that the presence of either of the two HLA risk alleles, *HLA-DQB1*03:03* and *HLA-C*07:04*, appeared to be predictive for response to cyclophosphamide (Rekeland et al., 2020). The original study by Lande et al. could not determine if the observed HLA associations were due to LD with another locus in the MHC, since it only investigated six of the classical HLA loci across this immunologically important, gene dense and polymorphic region.

The current study therefore aimed to fine map the HLA association signals in ME/CFS. To this end, we included genotyping of five additional classical HLA loci and 5,342 SNPs using 427 Norwegian ME/CFS cases and 480 controls. To our knowledge, this is the first study comprehensively investigating the association between the MHC and ME/CFS.

2. Material and methods

2.1. Study population

A total of 427 Norwegian ME/CFS patients and 480 healthy, ethnically matched controls were included in this study. All, but four, patients were diagnosed according to the 2003 Canadian Consensus Criteria (Carruthers et al., 2003). The total patient cohort was recruited from four separate inclusion groups; 1) the rituximab studies (Fluge, 2011; Fluge et al., 2019), 2) the cyclophosphamide study (Rekeland et al., 2020), 3) the CFS/ME biobank at Oslo University Hospital and 4) patients recruited via announcements in patient networks. Overlapping patients from the inclusion groups have been excluded as well as patients with self-reported non-Norwegian/Scandinavian ancestry. We included 38 patients from the cyclophosphamide study (two were excluded due to non-Norwegian ancestry), and tested the fine mapped HLA risk variants against the clinical response to cyclophosphamide.

2.2. SNP and HLA genotyping

SNPs were genotyped using the Infinium ImmunoArray-24 v2 BeadChip (iChip v2) from Illumina (Illumina, San Diego, USA) for the 427 cases, while the 480 healthy controls were genotyped using HumanImmuno-v1 BeadChip (iChip v1, Illumina), as described previously (Beecham, 2013). The data were merged using Plink v1.9. Subsequently, the SNPs from the MHC region on chromosome 6 (29.6–33.1 Mb, build 37) were extracted.

High resolution, targeted, next generation sequencing (NGS) of HLA class I and II genes was performed for the 427 ME/CFS patients by Lande et al. using NGSgo kit and NGSengine software (Gendx, Utrecht, the Netherlands) (Lande, 2020), with 3 samples being omitted due to lack of DNA and 4 additional samples being recruited to the current study. Five additional HLA loci (*DQA1, DPA1, DRB3, DRB4, DRB5*) were included in the current study. HLA genotyping for the 480 healthy controls was performed by NGS at Stanford University, USA (Creary, 2021), and not part of the paper by Lande (2020). Thus, all the 11 classical HLA genes *HLA-A, -B, -C, -DRB1, -DQB1, -DQA1, -DPB1, -DPA1, -DRB3, -DRB4* and *-DRB5* were included in this study. Due to 1% missing genotypes for some of the HLA genes, HLA imputation was conducted to add to the observed HLA genotypes. To assess the precision of the imputation results, we compared imputed vs. known HLA genotypes at 2-field resolution for the entire dataset which showed an overall concordance of around 99.8%.

2.3. Quality control and statistical analysis

Genotyping cluster-plots from iChip v2 were manually inspected and poor performing SNPs were excluded prior to further analyses. SNPs with low genotyping success rate (<99%) and minor allele frequency (<1%) were excluded. Likewise, SNPs deviating from Hardy Weinberg equilibrium ($p < 0.001$) in controls were excluded. Only successfully genotyped SNPs overlapping the iChip v1 and v2 were analyzed. A grand total of 5,342 SNPs passed quality control and were included in the analyses. For SNP imputation, the Michigan imputation server was used (Reference Panel: 1000G Phase 3v5 EUR, rsq filter R < 0.3, phasing via Eagle v2.4, Build 37) (Das et al., 2016), while HLA*IMP:03 was used for the subsequent HLA imputation (Motyer, 2016).

BIGDAWG package v 2.1 in R (Pappas et al., 2016) was used for initial amino acid analysis. Association analyses were otherwise performed using Unphased v 3.0.13 (Dudbridge, 2008). Amino acid calls to be included in association analyses using Unphased were based on codon information from the reference sequences in the IPD-IMGT/HLA database for the alleles and annotated as biallelic markers (present or not present). For LD plots, Haploview version 4.2 was used. Locuszoom was used to make plots visualizing association results with combined LD measures from publicly available 1000G EUR data (Pruim et al., 2010).

The Fisher's exact probability calculator, MedCalc was also used.

In this study, we used a significance threshold of $p \leq 0.001$, and a suggestive threshold of $p < 0.005$. This is primarily based on the previously observed significance levels of the HLA association in Lande (2020). At present time, genomic research in ME/CFS lacks large cohorts diagnosed according to the Canadian Consensus Criteria able to detect GWAS level significance.

2.4. Bioinformatic databases

The GTEx database was used to interrogate whether the associated SNPs could serve as expression quantitative trait loci (eQTL) by influencing the gene expression in various tissues. The data used for the analyses described in this manuscript were obtained from single tissue eQTL using the GTEx Portal on 05/20/21. The human protein atlas was used to obtain data on protein and gene expression for specific genes in various tissues (Uhlén, 2015).

3. Results

Among the genotyped SNPs across the extended MHC on chromosome 6 (29.6–33.1 Mb, build 37), 5,342 SNPs passed quality control and were used in the subsequent analyses. In addition, our dataset comprised of 222 classical 2-field HLA alleles from the 11 classical HLA genes, and 789 amino acids encoded by the HLA alleles were also tested.

3.1. Several ME/CFS associations across the MHC

Association analyses identified 60 associated markers ($p \leq 0.001$) across the MHC, including the HLA class I and II regions (Figure 1; Supplementary Table 1). None of the classical HLA loci or alleles reached the significance threshold of $p \leq 0.001$ (Supplementary Table 2). The class II gene, *HLA-DQB1*, was the most associated HLA locus displaying a p-value ($p = 0.004$) below the suggestive threshold ($p < 0.005$). Interestingly, HLA-DQB1 amino acid residues 57D and 57A did however reach the significance threshold of $p \leq 0.001$ ($p = 0.001$, both). Allele-wise, only the HLA class I allele *HLA-C*07:04* ($p = 0.004$) showed suggestive association. Five SNPs marked the distinct regions associated with ME/CFS ($p \leq 0.001$), i.e. from telomeric to centromeric: rs9366658 in the extended class I region (risk allele frequency (RAF) = 86.7% in cases vs 80.0% in controls; OR = 1.6; $p = 0.0003$), three SNPs in the HLA class I region namely rs388234 (RAF = 42.8% in cases vs 35.2% in controls; OR = 1.4; $p = 0.0008$), rs4711249, (RAF = 18.7% in cases vs 12.8% in controls; OR = 1.6; $p = 0.0005$) and rs3130477 (RAF = 90.6% in cases vs 85.6% in controls; OR = 1.6; $p = 0.0008$), and finally rs9275582 in the HLA class II region (RAF = 78.3% in cases vs 70.6% in controls; OR = 1.5; $p = 0.0002$).

Repeated analyses including the imputed SNP genotypes did not reveal any additional regions, but rather provided additional significant SNPs supporting the already observed association signals (Supplementary Fig. S1).

3.2. HLA class I and II regions are independently associated

Given the complex and extensive LD structure of the MHC, we next investigated the dependencies between the five tag SNPs of which one is located in the extended HLA class I region (rs9366658), three in the HLA class I region (rs388234, rs4711249, rs3130477), and one in the HLA class II region (rs9275582).

Overall, the five SNPs showed low correlation measured by $r^2 < 0.15$ (Figure 2 A and B), however, some haplotype structures were indicated from the higher D' values (Figure 2 C and D). Therefore, to further explore the dependency between the SNP associations, we performed conditional association analyses for each of the five SNPs separately. The class II SNP, rs9275582, remained significantly associated ($p < 0.0004$, Table 1) regardless of which of the other tag SNPs that was conditioned for, and likewise conditioning on rs9275582 did not affect the association signals seen for the four other SNPs ($p \leq 0.001$).

Contrary to this, the dependencies between the extended class I/class I SNPs (rs9366658, rs388234, rs4711249, and rs3130477) were less clear. The significance of the SNP associations were generally weakened after conditional analyses (Table 1). Interestingly, rs9366658 remained significantly associated ($p = 0.0009$) after conditioning on rs4711249, and likewise rs4711249 remained associated after conditioning on rs3130477 ($p = 0.0003$). Haplotype analyses between the telomeric SNP (rs9366658) in the extended class I region and the class I SNP (rs388234) showed significant association for those carrying either the risk alleles ($p = 0.0002$) or non-risk alleles ($p = 0.00004$) at both loci (Supplementary Table 3). However, the most common haplotype seen in the population (>45%) carried the risk allele at rs9366658 and the non-

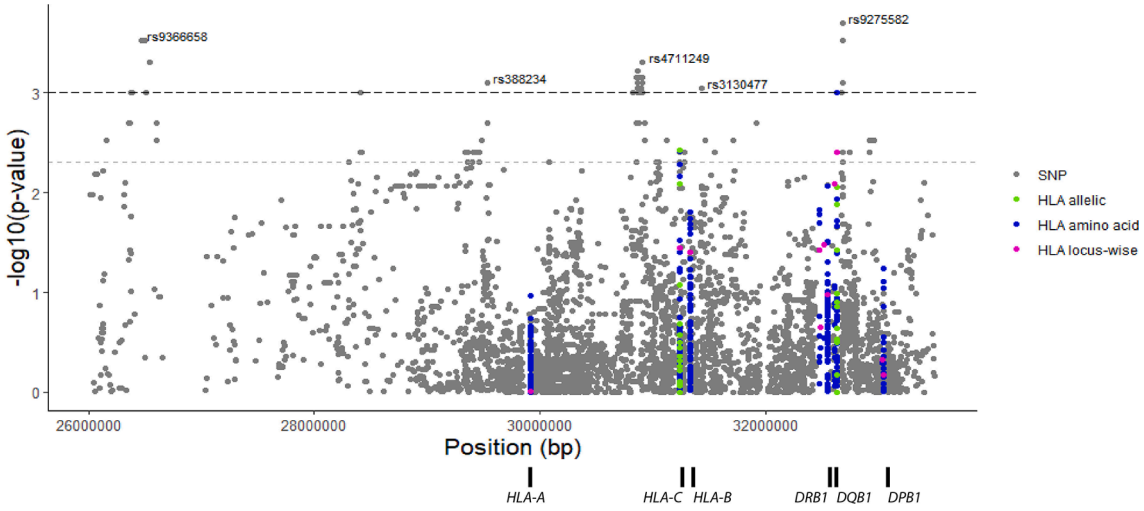


Fig. 1. Single marker association results across the major histocompatibility complex (MHC) in 427 Norwegian Myalgic encephalomyelitis/Chronic fatigue syndrome (ME/CFS) patients and 480 healthy controls. The statistical significance of the association analysis as -log10 of the P-value (y-axis), is plotted against the position on chromosome 6 shown in base pairs (bp, x-axis). The black, horizontal dashed line represents a significance threshold of $P = 0.001$, while the grey dashed line represents the suggestive threshold of $P = 0.005$. Association results are colored according to single nucleotide polymorphisms (SNPs), allelic *HLA-C* and *-DQB1* analyses (HLA allelic), HLA amino acid analyses and HLA locus-wise analyses. Positions are according to National Center for Biotechnology Information's build 37 (hg19). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

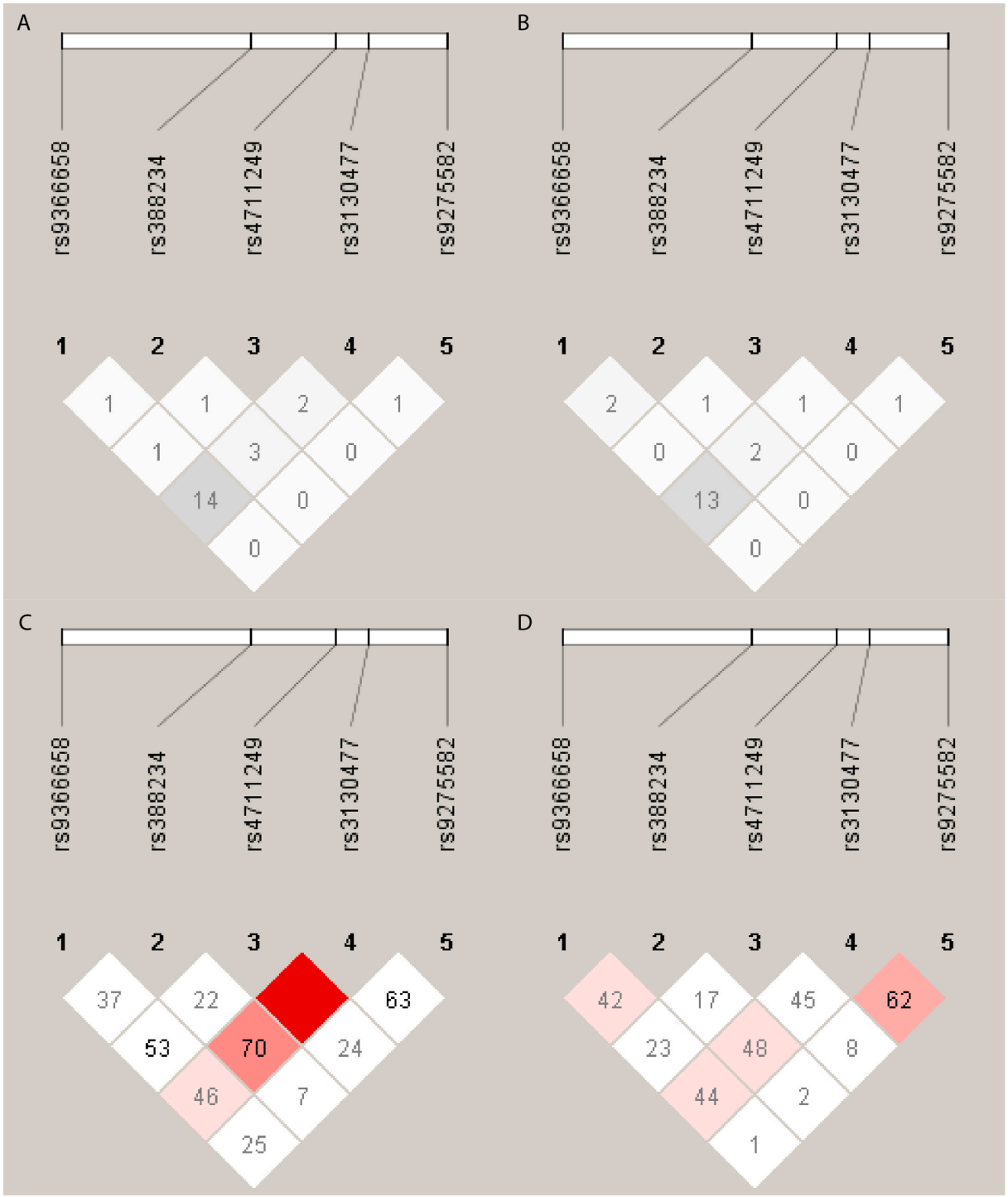


Fig. 2. Linkage disequilibrium (LD) plot for the five tag SNPs rs9366658, rs388234, rs4711249, rs3130477 and rs9275582 measured by r2 in cases (A) and controls (B), and by D' in cases (C) and controls (D) for 427 Norwegian ME/CFS patients and 480 healthy controls. Values shown in the figures are 100 × r^2 (A and B) and 100 × D' (C and D).

risk allele at rs388234, without being ME/CFS associated (p = 0.5). Only few haplotypes (<5%) carried the risk allele only at rs388234. Hence, rs9366658 appears to be secondarily associated and hitch-hiking on rs388234. A similar pattern was seen for the pairwise haplotype analysis between rs4711249 and rs3130477. The haplotype with risk alleles at both SNPs was positively associated (p = 0.0002), while the non-risk alleles were carried on a negatively associated haplotype (p = 0.001). The most common haplotype (>70%) carried the risk allele at rs3130477 but was not significantly associated (p = 0.5). Few haplotypes (<1%) only carried the risk allele at rs4711249.

Taken together, based on the SNP associations, we found three apparently independent association signals, best captured by the tag SNPs: rs388234 and rs4711249 in the HLA class I region and rs9275582 in the HLA class II region.

Table 1
Conditional analyses to explore the dependencies between the most significant SNPs tagging the five associated regions. BP is the position at chromosome 6 and P is P values for the unconditioned single marker association analysis.

			P values after conditioning on following SNP				
SNP	BP	P	rs9366658	rs388234	rs4711249	rs3130477	rs9275582
rs9366658	26469866	0.0003	1	0.002	0.0009	0.009	0.0002
rs388234	29533295	0.0008	0.02	1	0.003	0.005	0.001
rs4711249	30908266	0.0005	0.004	0.002	1	0.0003	0.001
rs3130477	31428920	0.0009	0.06	0.005	0.03	1	0.0006
rs9275582	32680070	0.0002	0.0004	0.0002	0.0003	0.00007	1

3.3. The relation between the ME/CFS associated tag SNPs and HLA-C and HLA-DQB1

We next addressed the relationship between these new SNP associations with our previously reported ME/CFS associations with *HLA-C* (C*07:04) and *HLA-DQB1* (DQB1*03:03). After conditioning on *HLA-C*, several SNP associations were still evident in the four initial association peaks (Figure 3A), including for rs388234 (p = 0.0005), telomeric of *HLA-A*. The SNP closest to *HLA-C*, rs4711249, did not remain associated after conditioning on *HLA-C* (p = 0.21), however, several other SNPs in the same LD region still showed strong association, e.g. rs7766094 (p = 0.0001), an intronic SNP in *VARS2*, annotated as splice donor variant.

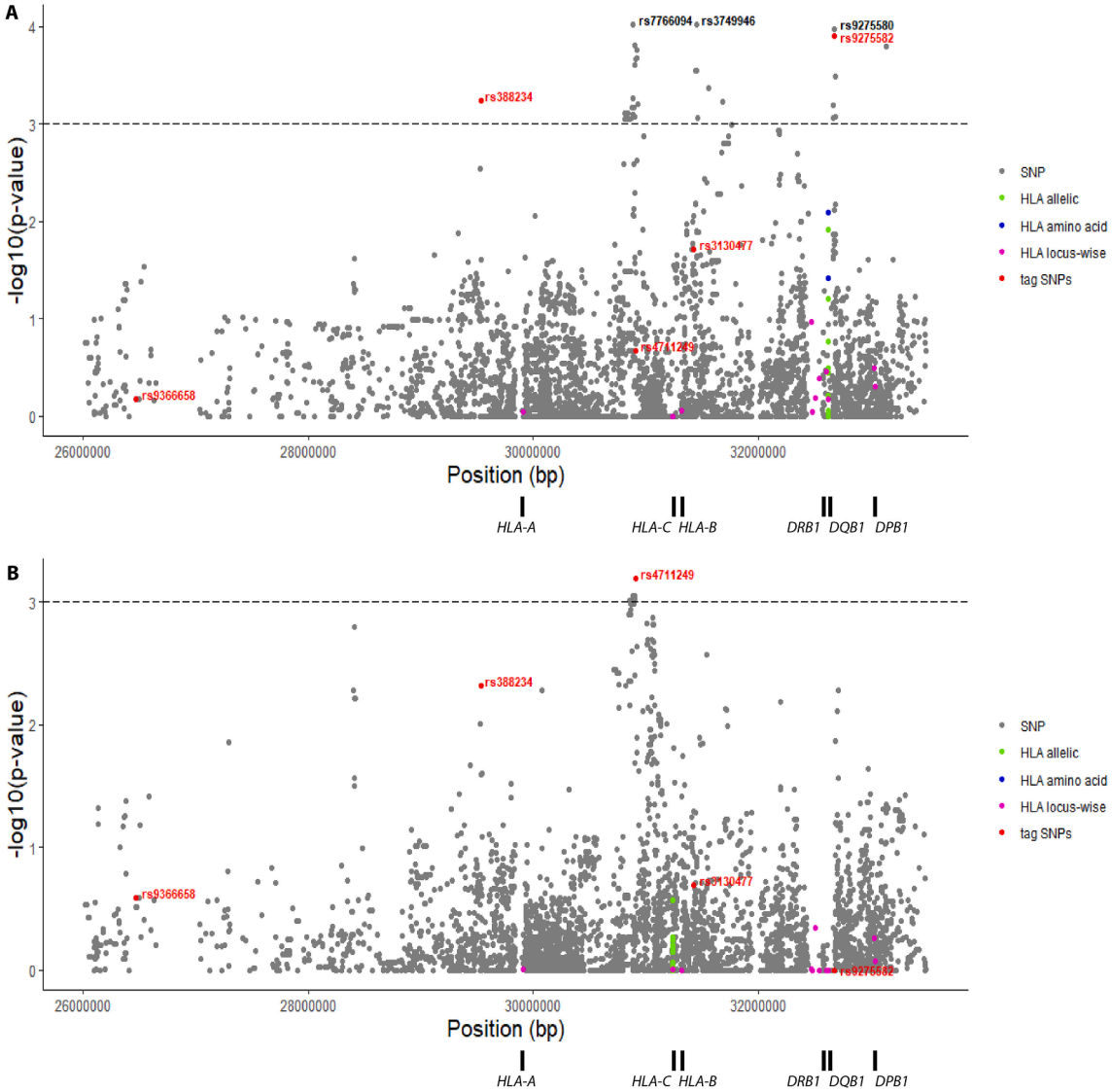


Fig. 3. Single marker association results within the major histocompatibility complex (MHC) in Myalgic encephalomyelitis/Chronic fatigue syndrome (ME/CFS) patients and healthy controls for analyses conditioning on A) *HLA-C* and B) *HLA-DQB1*. The statistical significance of the association analysis as -log10 of the P-value (y-axis), is plotted against the positions on chromosome 6 shown in base pairs (bp, x-axis). The black, horizontal dashed line represents a significance threshold of P = 0.001. Association results are colored according to single nucleotide polymorphisms (SNPs), allelic HLA-C and -DQB1 analyses (HLA allelic), HLA amino acid analyses, HLA locus-wise analyses and tag SNPs. Positions are according to National Center for Biotechnology Information's build 37 (hg19). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Interestingly, HLA-C*07:04 was almost always carried together with the risk allele at rs4711249 on a high-risk haplotype (3.3% in cases vs 1.3 % in controls; p = 0.001). However, the SNP risk allele was also carried on other *HLA-C* alleles (Supplementary Table 4), and overall haplotypes comprising the rs4711249 risk allele together with non-C*07:04 is still predisposing (15.4% in cases vs 11.5% in controls; p = 0.01).

Conditioning on *HLA-DQB1*, reduced the SNP associations below our significance threshold (Figure 3B), except for the class I associated SNP, rs4711249.

3.4. Closing in on the primary association in the HLA class I region

The SNP rs4711249 showed the most significant (p = 0.0005) association within the HLA class I region and is situated close to the *SFTA2* and *DPCR1* genes and almost 33 kb downstream of *HLA-C*. rs4711249 is in strong LD with several SNPs spanning the genes *DDR1*, *MIR4640*, *GTF2H4*, *VARS2*, *SFTA2* and *DPCR1* (Figure 4A). The expression levels of these genes differed significantly according to the genotype present at rs4711249 in multiple tissues (Supplementary Fig. 2). In general, the risk allele A was for most eQTLs associated with increased gene expression (i.e. *DDR1*, *VARS2*, and *SFTA2*), but also with decreased expression (i.e. *GTF2H4*, *SFTA2* and *DPCR1*). Interestingly, rs4711249 influences the gene expression in relevant tissues for ME/CFS, like nerves (*DDR1*, *SFTA2*, *DPCR1*), brain (*GTF2H4*), muscle (*GTF2H4*, *VARS2*) and blood (*DDR1*, *VARS2*).

The other associated SNP in the HLA class I region, rs388234 (p = 0.0008), is located in intron 18 of an alternative splice variant of the *GABBR1* gene (ENST00000355973.7). No strong LD with surrounding SNPs was observed (Figure 4B). No expression quantitative trait was observed for rs388234 and full length *GABBR1* (data not shown). The human protein atlas reveals that *GABBR1* is mainly expressed in brain, including in cerebellum, cerebral cortex and hippocampus (Supplementary Fig. 3). Notably, rs388234 lacks association support by surrounding SNPs, and the signal disappeared after conditioning on *HLA-DQB1* (Figure 3B).

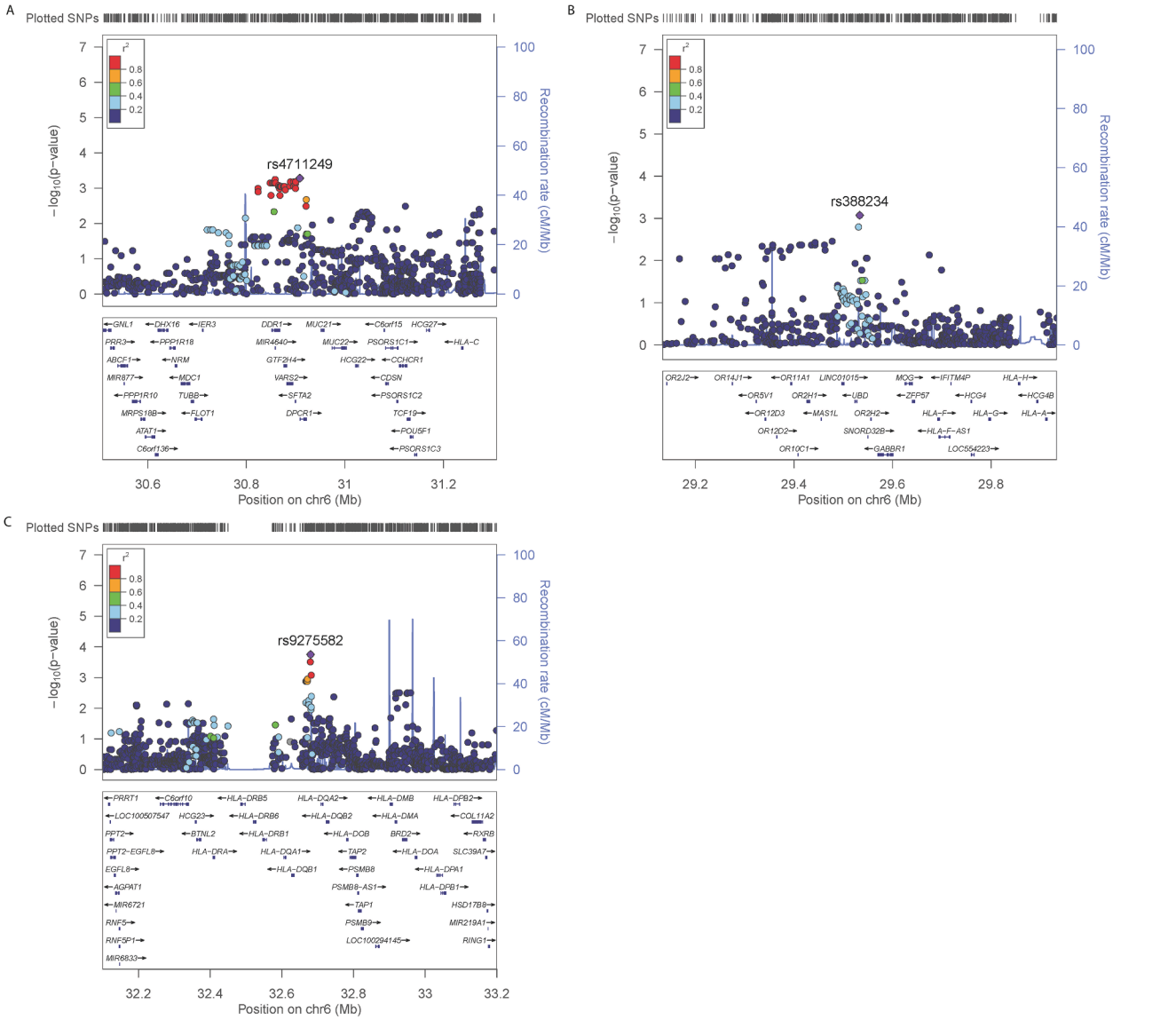


Fig. 4. The significance of the association signals from all SNPs, including their LD with the top SNP, across the disease susceptibility regions tagged by A) rs4711249, B) rs388234 and C) *HLA-DQB1*. The significance of the association signals, -log10P-value (y-axis), are plotted against the positions on chromosome 6 shown in base pairs (bp, x-axis). The colour of the dots is indicating r2 value. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.5. HLA-DQB1 could be primarily associated with disease risk by amino acid 57

In the HLA class II region, the rs9275582 was the most significantly associated SNP (p = 0.0002). Rs9275582 is located upstream of both *HLA-DQB1* and *-DQA2* and is in strong LD with surrounding SNPs (Figure 4C). Notably, the SNP density is very low across the *HLA –DR* and *-DQ* loci. The HLA-DQB1 association was best captured by the amino acid position 57 (57D: OR = 1.36, p = 0.001 and 57A: OR = 0.7, p = 0.001). Both amino acid 57A and 57D are present at multiple HLA-DQB1 alleles covered by our HLA genotyping (Table 2). Albeit none of the alleles show significant association per se, the alleles comprising 57D are generally increased among ME/CFS patients, while the 57A encoding alleles are increased among controls. After conditioning on residue 57A, the significance of rs9275582 dropped below our significance level (p = 0.004) and likewise after conditioning for 57D (p = 0.03). Furthermore, after adjusting for the whole *HLA-DQB1* locus, no residual association remained (p = 0.99, Figure 3B). Similarly, conditioning on rs9275582 abolished the association for the HLA-DQB1 amino acids 57A (p = 0.03) and 57D (p = 0.2). Accordingly, the rs9275582 risk allele shows strong positive LD (measured by D') with all common HLA-DQB1 alleles (>5%) encoding for the amino acid position 57D and strong negative LD (D'=-1) with the *HLA-DQB1*03:02* allele encoding the amino acid 57A (Supplementary Table 5).

3.6. Risk variants in relation to cyclophosphamide treatment response

We have previously reported that the *HLA* risk alleles *DQB1*03:03* and *C*07:04* are associated with response to cyclophosphamide treatment in ME/CFS patients (Rekeland et al., 2020). Notably, HLA-DQB1 57D was carried by almost all patients (94.7%) and hence was not useful in discriminating responders from non-responders. Since the most pronounced association in the class I region in the current study maps outside *HLA-C*, we investigated the rs4711249 risk allele against cyclophosphamide treatment response in 38 individuals. In total 39.5% (15 out of 38) of the patients from the cyclophosphamide study carried either the rs4711249 risk allele and/or *DQB1*03:03*. Interestingly, 61.9 % (13 out of 21) of responders to cyclophosphamide carried either the rs4711249 risk allele and/or *DQB1*03:03* vs 11.8% (2 out of 17) of non-responders. A similar trend was observed for patients having *HLA-C*07:04* and/or *DQB1*03:03* (81.8 % or 9 out of 11 patients carrying the alleles vs. 44.4% or 12 out of 27 non-carriers were responders).

Table 2
Association between ME/CFS and amino acid D or A in position 57 of DQB1, and their respective HLA-DQB1 alleles.

	Frequency in cases, %	Frequency in controls, %	P
Residue 57 aspartic acid (57D)	53.6	46.0	0.001
Alleles encoding 57D:			
*DQB1*03:01*	15.0	13.3	0.3
*DQB1*03:03*	6.4	4.7	0.1
*DQB1*03:09*	0	0.1	1
*DQB1*03:19*	0.1	0	0.5
*DQB1*04:02*	4.6	4.2	0.7
*DQB1*05:03*	2.8	1.8	0.1
*DQB1*06:01*	0	0.5	0.06
*DQB1*06:02*	17.0	14.9	0.2
*DQB1*06:03*	7.7	6.5	0.3
*DQB1*06:16*	0	0.1	1
Residue 57 alanine (57A)	28.8	35.9	0.001
Alleles encoding 57A:			
*DQB1*02:01*	11.4	13.8	0.1
*DQB1*02:02*	5.3	5.7	0.7
*DQB1*03:02*	12.1	16.4	0.009
*DQB1*03:04*	0.1	0.1	1

4. Discussion

By combining high-quality HLA genotyping and SNP genotyping data in unconditional and conditional analyses, we showed that there are independent associations with HLA class I and class II region in ME/CFS.

Association with HLA genes within the MHC is a hallmark for immune-mediated diseases. However, to pinpoint the primary associations can be challenging due to the complex underlying LD and haplotype structures. In the recent study by our group, we found ME/CFS associations with the *HLA-C*07:04* and *HLA-DQB1*03:03* alleles when investigating 6 of the 11 classical HLA genes (Lande, 2020). In the current study, we investigated all 11 classical HLA genes in combination with 5,342 SNPs.

We initially identified several potentially associated regions across the extended MHC tagged by five SNPs. However, several of these appeared to be secondary due to LD and extended haplotype structures, leaving two associated regions in the HLA class I region and one in the HLA class II region. Furthermore, after scrutinizing the relationship between the association signals through LD, conditional and haplotype analyses, we observed that HLA class I and II were independently associated with ME/CFS.

The most pronounced association in the HLA class I region was with rs4711249. This SNP is most likely not reflecting an association with *HLA-C*. Interestingly, the association between HLA risk alleles for ME/CFS and cyclophosphamide response was maintained even if we replaced the *HLA-C*07:04* allele with the risk allele at rs4711249. However, due to complex LD patterns, it is difficult to conclude that the association maps outside the classical HLA locus, however, the numerous associated SNPs tagged by rs4711249 spans several genes (*DDR1*, *GTF2H4*, *VARS2*, *SFTA2* and *DPCR1*) being candidates for involvement in ME/CFS. Interestingly, rs4711249 genotype was associated with gene expression differences of these surrounding genes in both muscle, brain, blood and nervous tissue. Hence, the risk variant could be a regulatory variant that provide susceptibility to ME/CFS through altering gene expression levels. The *VARS2* gene is implicated in several mitochondrial diseases with causative mutations reported for mitochondrial encephalopathy and putative pathogenic variants have been found for mitochondrial respiratory complex deficiencies (Diodato et al., 2014; Taylor et al., 2014). This is interesting given the clinical manifestations of ME/CFS and the proposed, but not yet established, alterations of mitochondrial respiratory function related to ME/CFS (Holden, 2020).

The other HLA class I association, rs388234, is located in an alternative intron of a splice variant of the *GABBR1* gene. Unlike rs4711249, we did not observe eQTLs for rs388234 and full length *GABBR1*. However, the human protein atlas shows that *GABBR1* is mainly expressed in the brain. The *GABBR1* gene encodes a receptor for gamma-aminobutyric acid (GABA), which is the main inhibitory neurotransmitter in the mammalian central nervous system. Changes in this gene may underlie brain disorders and may be involved in regulating proliferation of hematopoietic stem cells (Xi et al., 2011; Elujoba-Bridenstine, et al., 2019). The robustness of the association with rs388234 is weak, as the signal is not supported by any surrounding SNPs, and the rs388234 association disappeared after conditioning on *HLA-DQB1*.

Additionally, for the HLA class II region, our analyses point towards the main association signal being within the HLA-DQ locus. However, we could not determine whether rs9275582 or *HLA-DQB1* represent the primary association due to strong LD.

Lande et al. found *HLA-DQB1*03:03* to be associated with risk of ME/CFS, and we observed the same trend, albeit not significant, for the same *HLA-DQB1* allele (6.4% vs 4.7%, OR = 1.66) with our smaller and distinct control data set. However, it seems likely that if *HLA-DQB1* is the main class II risk locus, it is not conferred by one specific allele, but rather an amino acid position as we observe significant association with amino acid residue 57D which is coded by *HLA-DQB1*03:03* and several

other alleles. In fact, none of the alleles coding for amino acid residue 57D showed any significant associations at the allelic level, thus illustrating the need for multi-level analysis of HLA. The HLA-DQB1 residue 57 has been implicated in immune-mediated disease previously, i.e. in type 1 diabetes where 57A confers susceptibility while 57D provides protection (Gough and Simmonds, 2007; Holoshitz, 2010). This amino acid is a crucial residue in the peptide-binding groove of HLA-DQB1 involved in antigen presentation and T cell receptor interaction. The presence of different amino acids at this residue could possibly modulate the stability of the molecule thereby influencing the binding of peptides (Gough and Simmonds, 2007).

In most diseases, it has been difficult to identify the main genetic determinants in the MHC region (Fiorillo et al., 2017). Strong LD, the density of immune related genes, and the additive effect of multiple HLA loci are all key factors. Unfortunately, few HLA studies have been undertaken in ME/CFS, and overall, they lack statistical power, strict diagnostic criteria for ME/CFS, high resolution HLA genotypes and fine mapping using markers spanning the MHC. Hence, our work marks, to the best of our knowledge, the first thorough and rigorous HLA analysis and fine mapping in ME/CFS. The strength of our study is the dense genetic map including deep sequencing of 11 classical HLA genes and ME/CFS patients diagnosed according to the strict Canadian criteria. Nevertheless, it must be acknowledged that we are, despite having 4–10 times as many patients as previous studies, still lacking statistical power. Consequently, we used a significance threshold of $p \leq 0.001$ without correcting for multiple testing. In established autoimmune diseases, albeit the initials studies only comprised of hundreds of patients, the genetic determinants in the MHC region have now been fine mapped in several thousand patients. In addition, we have, to the best of our ability, tried to incorporate homogeneously diagnosed patients as well as the use of best practice and QC in data analysis. The statistical power issue will hopefully be addressed in future studies given the growing cooperation between groups, and the on-going collection of DNA from ME/CFS patients, like the DecodeME project (www.decodeme.org.uk). This will enable further studies of larger and more powerful cohorts, which are warranted to firmly establish the involvement of the MHC and HLA genes in ME/CFS.

In addition, the COVID-19 pandemic might increase the number of people who develop ME/CFS given the substantial overlap between ME/CFS and post COVID syndrome (Komaroff and Lipkin, 2021). The commonalties between the two diseases could possibly reveal similar HLA predispositions which might lead to better understanding of both pathologies.

In conclusion, we report, to the best of our knowledge, the first HLA fine mapping study in ME/CFS, in an attempt to address a putative immunologic component of the ME/CFS pathogenesis. We observed independent association signals from the HLA class I and class II regions, which encourage future genetic studies of the MHC in ME/CFS.

Funding

Funding was provided by the Kavli Trust and the Research Council of Norway. This project was supported in part by grant U19NS095774 from the U.S. National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not necessarily reflect the views of the NIH. The funders did not influence the research or the manuscript in any way.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are very grateful to the patients who participated in the study, and we thank the Norwegian Bone Marrow Donor Registry for access to controls. Thank you to Kari Sørland, Kine Alme and Kristin Risa at Department of Oncology and Medical Physics, Haukeland University Hospital, Wenche Kristiansen and Hilde Haukeland at the CFS/ME Biobank, Oslo University Hospital, Siri T. Flåm at Department of Medical Genetics, Oslo University Hospital, and employees at the Norwegian Sequencing Centre, University of Oslo and Oslo University Hospital and at the Genomics Core Facility, Oslo University hospital, Norway for all their help. The Genomics Core Facility, Oslo University Hospital and University of Oslo, performed the SNP genotyping. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbi.2021.08.219.

References

Albright, F., Light, K., Light, A., Bateman, L., Cannon-Albright, L.A., 2011. Evidence for a heritable predisposition to Chronic Fatigue Syndrome. BMC Neurol. 11 (1) https://doi.org/10.1186/1471-2377-11-62.

Beecham, A.H., et al., 2013. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat. Genet. 45 (11), 1353–1360.

Buchwald, D., et al., 2001. A Twin Study of Chronic Fatigue. Psychosom. Med. 63 (6).

Carlo-Stella, N., Bozzini, S., De Silvestri, A., Sbarsi, I., Pizzochero, C., Lorusso, L., Martinetti, M., Cuccia, M., 2009. Molecular Study of Receptor for Advanced Glycation Endproduct Gene Promoter and Identification of Specific HLA Haplotypes Possibly Involved in Chronic Fatigue Syndrome. Int.. J. Immunopathol. Pharmacol. 22 (3), 745–754.

Carruthers, B.M., Jain, A.K., De Meirleir, K.L., Peterson, D.L., Klimas, N.G., Lerner, A.M., Bested, A.C., Flor-Henry, P., Joshi, P., Powles, A.C.P., Sherkey, J.A., van de Sande, M.I., 2003. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. J. Chronic Fatigue Syndrome 11 (1), 7–115.

Creary, L.E., et al., 2021. Next-Generation Sequencing Identifies Extended HLA Class I and II Haplotypes Associated With Early-Onset and Late-Onset Myasthenia Gravis in Italian, Norwegian, and Swedish Populations. Front. Immunol. 12 (1754).

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E. Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P.-R., Iacono, W.G., Swaroop, A., Scott, L.J., Cucca, F., Kronenberg, F., Boehnke, M., Abecasis, G.R., Fuchsberger, C., 2016. Next-generation genotype imputation service and methods. Nat. Genet. 48 (10), 1284–1287.

Diodato, D., Melchionda, L., Haack, T.B., Dallabona, C., Baruffini, E., Donnini, C., Granata, T., Ragona, F., Balestri, P., Margollicci, M., Lamantea, E., Nasca, A., Powell, C.A., Minczuk, M., Strom, T.M., Meitinger, T., Prokisch, H., Lamperti, C., Zeviani, M., Ghezzi, D., 2014. VARS2 and TARS2 mutations in patients with mitochondrial encephalomyopathies. Hum. Mutat. 35 (8), 983–989.

Dudbridge, F., 2008. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. Hum. Hered. 66 (2), 87–98.

Elujoba-Bridenstine, A., et al., 2019. The Neurotransmitter Receptor Gabbr1 Regulates Proliferation and Function of Hematopoietic Stem and Progenitor Cells. Blood 134 (Supplement_1), 3707.

Feiring, B., Laake, I., Bakken, I.J., Greve-Isdahl, M., Wyller, V.B., Håberg, S.E., Magnus, P., Trogstad, L., 2017. HPV vaccination and risk of chronic fatigue syndrome/myalgic encephalomyelitis: A nationwide register-based study from Norway. Vaccine 35 (33), 4203–4212.

Fiorillo, M.T., Paladini, F., Tedeschi, V., Sorrentino, R., 2017. HLA Class I or Class II and Disease Association: Catch the Difference If You Can. Front. Immunol. 8 https://doi.org/10.3389/fimmu.2017.01475.

Fluge, Ø., et al., 2011. Benefit from B-lymphocyte depletion using the anti-CD20 antibody rituximab in chronic fatigue syndrome. A double-blind and placebo-controlled study. PLoS One 6 (10), e26358.

Fluge, Ø., Rekeland, I.G., Lien, K., Thürmer, H., Borchgrevink, P.C., Schäfer, C., Sørland, K., Aßmus, J., Ktoridou-Valen, I., Herder, I., Gotaas, M.E., Kvammen, Ø., Baranowska, K.A., Bohnen, L.M.L.J., Martinsen, S.S., Lonar, A.E., Solvang, A.-E., Gya, A.E.S., Bruland, O., Risa, K., Alme, K., Dahl, O., Mella, O., 2019. B-Lymphocyte Depletion in Patients With Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. Ann. Intern. Med. 170 (9), 585. https://doi.org/10.7326/M18-1451.

Gough, S.C., Simmonds, M.J., 2007. The HLA Region and Autoimmune Disease: Associations and Mechanisms of Action. Curr. Genomics 8 (7), 453–465.

Helbig, K.J., Heatley, S.L., Harris, R.J., Mullighan, C.G., Bardy, P.G., Marmion, B.P., 2003. Variation in immune response genes and chronic Q fever. Concepts: preliminary test with post-Q fever fatigue syndrome. Genes Immun. 4 (1), 82–85.

Holden, S., et al., 2020. A systematic review of mitochondrial abnormalities in myalgic encephalomyelitis/chronic fatigue syndrome/systemic exertion intolerance disease. J. Transl. Med. 18 (1), 290.
Holoshitz, J., 2010. The rheumatoid arthritis HLA-DRB1 shared epitope. Curr. Opin. Rheumatol. 22 (3), 293–298.
Keller, R.H., et al., 1994. Association Between HLA Class II Antigens and the Chronic Fatigue Immune Dysfunction Syndrome. Clin. Infect. Dis. 18 (Supplement_1), S154–S156.
Komaroff, A.L., Lipkin, W.I., 2021. Insights from myalgic encephalomyelitis/chronic fatigue syndrome may help unravel the pathogenesis of postacute COVID-19 syndrome. Trends Mol. Med. p. S1471–4914(21), 00134–9.
Lande, A., et al., 2020. Human Leukocyte Antigen alleles associated with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). Sci. Rep. 10 (1), 5267.
Lawrie, S.M., Pelosi, A.J., 1995. Chronic Fatigue Syndrome in the Community Prevalence and Associations. Br. J. Psychiatry 166 (6), 793–797.
Motyer, A., et al., 2016. Practical Use of Methods for Imputation of HLA Alleles from SNP. Genotype Data. bioRxiv, 091009.
Ortega-Hernandez, O.-D., et al., 2009. Autoantibodies, Polymorphisms in the Serotonin Pathway, and Human Leukocyte Antigen Class II Alleles in Chronic Fatigue Syndrome. Ann. New York Acad. Sci. 1173 (1), 589–599.
Pappas, D.J., Marin, W., Hollenbach, J.A., Mack, S.J., 2016. Bridging ImmunoGenomic Data Analysis Workflow Gaps (BIGDAWG): An integrated case-control analysis pipeline. Hum. Immunol. 77 (3), 283–287.
Pasi, A., et al., 2011. Excess of activating killer cell immunoglobulin-like receptors and lack of HLA-Bw4 ligands: A two-edged weapon in chronic fatigue syndrome. Mol. Med. Rep. 4 (3), 535–540.
Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., Willer, C.J., 2010. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26 (18), 2336–2337.
Rasa, S., Nora-Krukle, Z., Henning, N., Eliassen, E., Shikova, E., Harrer, T., Scheibenbogen, C., Murovska, M., Prusty, B.K., 2018. Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). J. Transl. Med. 16 (1) https://doi.org/10.1186/s12967-018-1644-y.
Rekeland, I.G., Fosså, A., Lande, A., Ktoridou-Valen, I., Sørland, K., Holsen, M., Tronstad, K.J., Risa, K., Alme, K., Viken, M.K., Lie, B.A., Dahl, O., Mella, O., Fluge, Ø., 2020. Intravenous Cyclophosphamide in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. An Open-Label Phase II Study. Front. Med. 7.
Sepúlveda, N., Carneiro, J., Lacerda, E., Nacul, L., 2019. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome as a Hyper-Regulated Immune System Driven by an Interplay Between Regulatory T Cells and Chronic Human Herpesvirus Infections. Front. Immunol. 10 https://doi.org/10.3389/fimmu.2019.0268410.3389/fimmu.2019.02684.s001.
Smith, J., et al., 2005. Association of chronic fatigue syndrome with human leucocyte antigen class II alleles. J. Clin. Pathol. 58 (8), 860–863.
Sotzny, F., Blanco, J., Capelli, E., Castro-Marrero, J., Steiner, S., Murovska, M., Scheibenbogen, C., 2018. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome – Evidence for an autoimmune disease. Autoimmun. Rev. 17 (6), 601–609.
Spitzer, A.R., Broadman, M., 2010. A Retrospective Review of the Sleep Characteristics in Patients with Chronic Fatigue Syndrome and Fibromyalgia. Pain Practice 10 (4), 294–300.
Taylor, R.W., Pyle, A., Griffin, H., Blakely, E.L., Duff, J., He, L., Smertenko, T., Alston, C. L., Neeve, V.C., Best, A., Yarham, J.W., Kirschner, J., Schara, U., Talim, B., Topaloglu, H., Baric, I., Holinski-Feder, E., Abicht, A., Czermin, B., Kleinle, S., Morris, A.A.M., Vassallo, G., Gorman, G.S., Ramesh, V., Turnbull, D.M., Santibanez-Koref, M., McFarland, R., Horvath, R., Chinnery, P.F., 2014. Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies. JAMA 312 (1), 68. https://doi.org/10.1001/jama.2014.7184.
Thorsby, E., Lie, B.A., 2005. HLA associated genetic predisposition to autoimmune diseases: Genes involved and possible mechanisms. Transpl. Immunol. 14 (3-4), 175–182.
Trowsdale, J., Knight, J.C., 2013. Major histocompatibility complex genomics and human disease. Annu. Rev. Genomics Hum. Genet. 14 (1), 301–323.
Uhlén, M., et al., 2015. Proteomics. Tissue-based map of the human proteome. Science 347 (6220), 1260419.
Underhill, J.A., et al., 2001. Lack of association between HLA genotype and chronic fatigue syndrome. Eur. J. Immunogenet. 28 (3), 425–428.
Xi, B., et al., 2011. GABBR1 gene polymorphism(G1465A)isassociated with temporal lobe epilepsy. Epilepsy Res. 96 (1-2), 58–63.

Supplementary Tables

Supplementary table 1. All markers across the MHC significantly (p≤0.001) associated with ME/CFS

SNP	BP	Risk allele	RAF in cases	RAF in controls	OR (95% CI)	P
rs9275582	32680070	G	0.783	0.706	1.50 (1.21-1.86)	0.0002
rs9275580	32679462	A	0.779	0.704	1.48 (1.19-1.83)	0.0003
rs9366658	26469866	G	0.868	0.805	1.59 (1.23-2.05)	0.0003
rs3799380	26467182	A	0.863	0.800	1.58 (1.23-2.02)	0.0003
rs6456728	26477779	G	0.863	0.800	1.58 (1.23-2.02)	0.0003
rs2093169	26495099	G	0.868	0.805	1.59 (1.23-2.05)	0.0003
rs7756567	26481642	A	0.868	0.805	1.58 (1.23-2.04)	0.0003
rs7773938	26474044	G	0.868	0.805	1.58 (1.23-2.04)	0.0003
rs9358944	26469875	A	0.868	0.805	1.58 (1.23-2.04)	0.0003
rs9358945	26472114	A	0.868	0.805	1.58 (1.23-2.04)	0.0003
rs9358946	26478927	G	0.868	0.805	1.58 (1.23-2.04)	0.0003
rs4711249	30908266	A	0.187	0.128	1.57 (1.22-2.03)	0.0005
rs6920256	26537801	G	0.915	0.864	1.69 (1.24-2.28)	0.0005
rs7743661	30858254	G	0.199	0.139	1.55 (1.21-1.98)	0.0006
rs1049622	30858857	A	0.198	0.139	1.53 (1.20-1.97)	0.0007
rs2229933	30857072	C	0.198	0.139	1.53 (1.20-1.97)	0.0007
rs2074511	30889389	G	0.186	0.128	1.56 (1.21-2.01)	0.0007
rs2286655	30899746	A	0.186	0.128	1.56 (1.21-2.01)	0.0007
rs3873332	30895990	G	0.186	0.128	1.56 (1.21-2.01)	0.0007
rs4618569	30855251	A	0.198	0.139	1.53 (1.20-1.97)	0.0007
rs7756521	30848253	G	0.198	0.139	1.53 (1.20-1.97)	0.0007
rs9468842	30852747	A	0.198	0.139	1.53 (1.20-1.97)	0.0007
rs4713402	30899233	A	0.185	0.128	1.54(1.20-2.00)	0.0007
rs3130477	31428920	A	0.906	0.856	1.63(1.21-2.19)	0.0009
rs3998158	32681992	A	0.288	0.219	1.43(1.16-1.68)	0.0008

rs388234	29533295	G	0.429	0.352	1.38(1.14-1.67)	0.0008
rs11542628	30876034	A	0.198	0.140	1.52(1.18-1.95)	0.0008
rs1362123	30876996	G	0.198	0.140	1.52(1.18-1.95)	0.0008
rs2239517	30865115	A	0.198	0.140	1.52(1.18-1.95)	0.0008
rs2284175	30875145	G	0.198	0.140	1.52(1.18-1.95)	0.0008
rs2284176	30875622	G	0.198	0.140	1.52(1.18-1.95)	0.0008
rs2894055	30868628	G	0.198	0.140	1.52(1.18-1.95)	0.0008
rs3869086	30870168	A	0.198	0.140	1.52(1.18-1.95)	0.0008
rs8408	30867666	G	0.198	0.140	1.52(1.18-1.95)	0.0008
rs916920	30877202	G	0.198	0.140	1.52(1.18-1.95)	0.0008
rs9468845	30869593	A	0.198	0.140	1.52(1.18-1.95)	0.0008
rs9468846	30870763	G	0.198	0.140	1.52(1.18-1.95)	0.0008
rs2285319	30887972	A	0.192	0.134	1.53(1.20-1.97)	0.0008
rs3873334	30896147	A	0.192	0.134	1.53(1.20-1.97)	0.0008
rs4711247	30895680	A	0.192	0.134	1.53(1.20-1.97)	0.0008
rs2239518	30865725	A	0.185	0.128	1.54(1.20-2.00)	0.0008
rs3799383	26510748	G	0.912	0.863	1.64(1.21-2.22)	0.0009
rs12176317	26372786	A	0.915	0.868	1.64(1.20-2.22)	0.001
rs1977	26377546	A	0.915	0.868	1.64(1.20-2.22)	0.001
rs2073529	26377546	A	0.915	0.868	1.64(1.20-2.22)	0.001
rs9393713	26377546	G	0.915	0.868	1.64(1.20-2.22)	0.001
rs9275332	32666943	G	0.733	0.663	1.39(1.13-1.70)	0.001
rs9275371	32668296	A	0.733	0.663	1.39(1.13-1.70)	0.001
rs9275393	32669439	G	0.733	0.663	1.39(1.13-1.70)	0.001
rs9275407	32670037	C	0.733	0.663	1.39(1.13-1.70)	0.001
rs9275428	32670978	A	0.733	0.663	1.39(1.13-1.70)	0.001
rs9275439	32671521	A	0.733	0.663	1.39(1.13-1.70)	0.001
rs2531804	28411303	A	0.690	0.617	1.38(1.13-1.38)	0.001
rs2074512	30878919	G	0.196	0.139	1.51(1.17-1.93)	0.001
rs9295928	30823630	G	0.177	0.123	1.53(1.18-1.98)	0.001

rs3916232	30823366	T	0.178	0.123	1.54(1.20-2.00)	0.001
rs2286656	30899571	A	0.179	0.124	1.54(1.20-2.00)	0.001
rs1049628	30867106	A	0.198	0.140	1.51(1.18-1.94)	0.001
DQB1 57D	32632687-32632689		0.536	0.460	1.36(1.27-1.63)	0.001
DQB1 57A*	32632687-32632689		0.288	0.359	0.72(0.59-0.87)	0.001

Five tag SNPs are marked in bold, BP is nucleotide position at chromosome 6, RAF is risk allele frequency, OR is odds ratio, 95% CI is the 95% confidence interval and P is P value, DQB1 57D is HLA-DQB1 amino acid residue 57 aspartic acid (D), DQB1 57A is HLA-DQB1 amino acid residue 57 alanine (A).

* DQB1 57A is a negatively associated amino acid.

Supplementary table 2. Association analyses of classical HLA loci (all global p-values) and HLA alleles showing suggestive association.

Locus	**P**
HLA-A	0.98
HLA-C	0.04
HLA-B	0.04
HLA-DRB3	0.04
HLA-DRB5	0.22
HLA-DRB4	0.03
HLA-DRB1	0.11
HLA-DQA1	0.008
HLA-DQB1	0.004
HLA-DPA1	0.47
HLA-DPB1	0.67
HLA-C*07:04	0.004

Supplementary table 3. Pairwise haplotype association analyses between the neighbouring tag SNPs in extended class I/class I tag SNPs.

rs9366658-rs388234	Ca-Freq	Co-Freq	OR	P
A-**G**	0.036	0.040	reference	0.5
A-A	0.097	0.155	0.69	3.6e-005
G-**G**	0.393	0.312	1.39	0.0002
G-A	0.475	0.493	1.1	0.4
rs388234-rs4711249	Ca-Freq	Co-Freq	OR	P
G-**A**	0.105	0.059	1.4	0.0002
G-G	0.324	0.293	0.62	0.1
A-**A**	0.083	0.069	reference	0.2
A-G	0.489	0.579	0.47	4.4e-005
rs4711249-rs3130477	Ca-Freq	Co-Freq	OR	P
A-G	0	0.010	-	2.8e-008
A-**A**	0.189	0.118	1.6	0.0002
G-G	0.093	0.134	0.7	0.001
G-**A**	0.718	0.738	reference	0.5

Risk alleles from single SNP analyses are bold. Ca-Freq is frequency in cases, Co-Freq is frequency in controls, OR is odds ratio, P is p-value

Supplementary table 4. Association analyses for haplotypes of rs4711249 and *HLA-C*:

Haplotype	Ca-Freq	Co-Freq	OR	P
A-HLA-C*07:04	0.034	0.013	1.6	0.001
A-HLA-C (all other alleles)	0.154	0.115	1.4	0.01
G-HLA-C*07:04	0.005	0.003	1.5	0.5
G-HLA-C (all other alleles)	0.807	0.868	0.5	0.0004

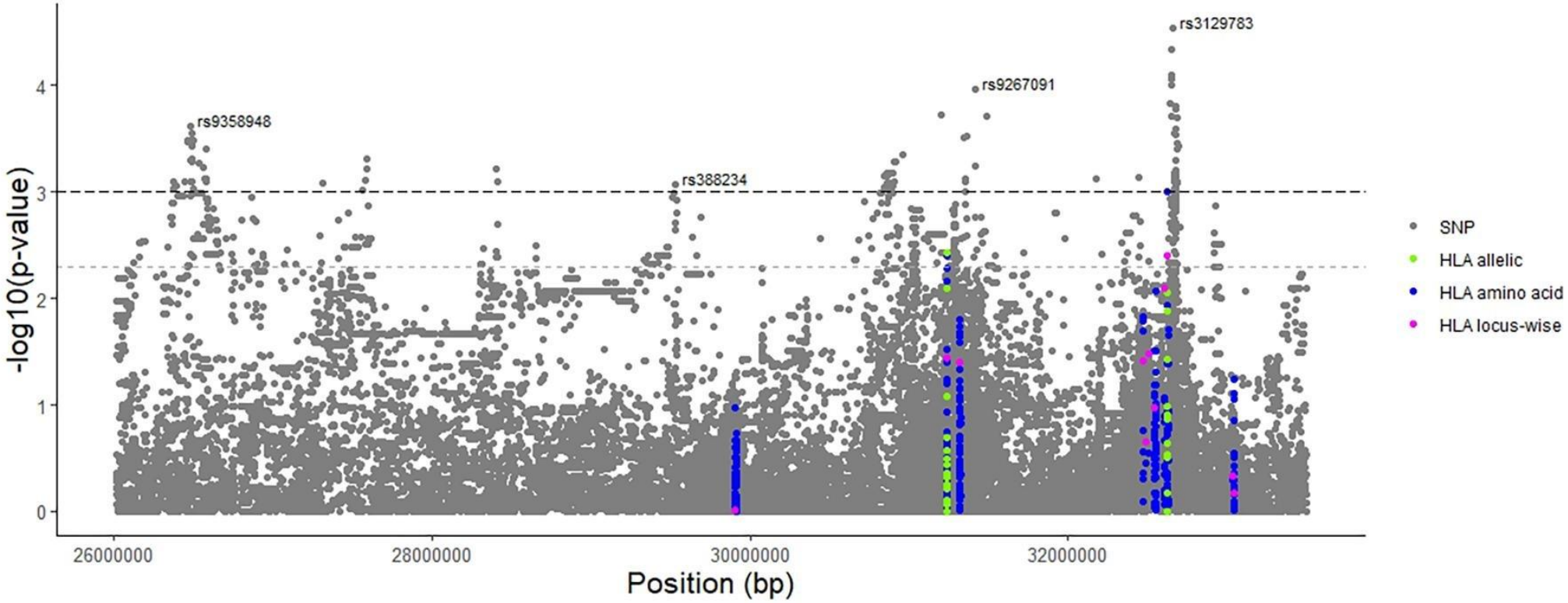
Ca-Freq is frequency in cases, Co-Freq is frequency in controls, OR is odds ratio, P is p-value

Supplementary table 5. LD measures between *HLA-DQB1* alleles and rs9275582 risk allele G

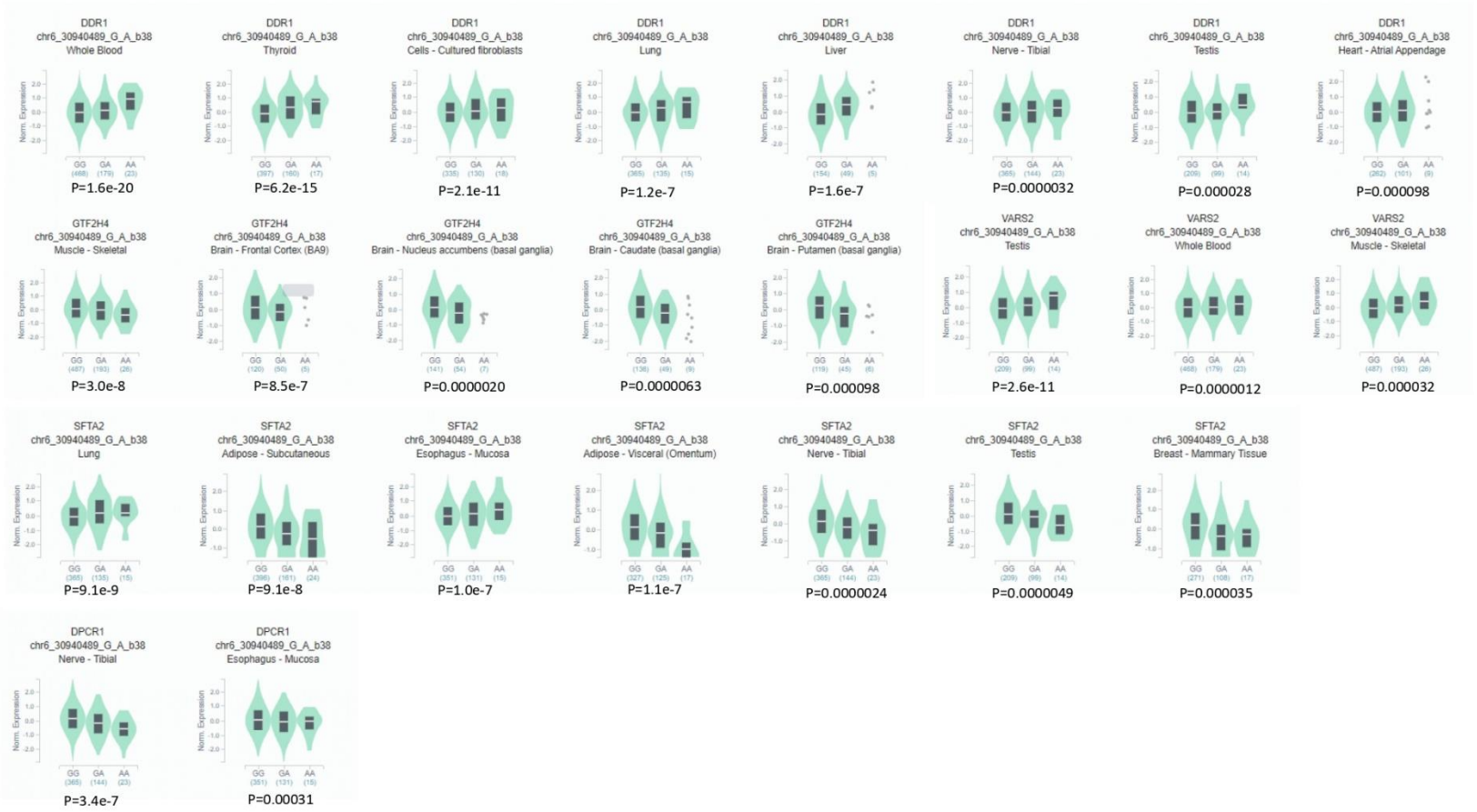
	ca-D'	co-D'	ca-r^2	co-r^2
DQB1*02:01	1	1	0.04	0.07
DQB1*02:02	1	1	0.02	0.03
DQB1*03:01	1	1	0.05	0.06
DQB1*03:02	-1	-1	0.50	0.47
DQB1*03:03	1	1	0.02	0.02
DQB1*03:04	1	1	0.0003	0.0004
DQB1*03:09	0	1	0	0.0004
DQB1*03:19	-1	0	0.0004	0
DQB1*04:02	1	0.85	0.01	0.01
DQB1*05:01	-0.98	-1	0.37	0.35
DQB1*05:02	1	1	0.0003	0.001
DQB1*05:03	1	1	0.008	0.007
DQB1*05:04	1	1	0.002	0.005
DQB1*05:84	0	-1	0	0.005
DQB1*06:01	0	1	0	0.002
DQB1*06:02	1	1	0.06	0.07
DQB1*06:03	1	1	0.02	0.03
DQB1*06:04	1	1	0.02	0.02
DQB1*06:09	-0.12	0.52	0.0002	0.0005
DQB1*06:16	0	1	0	0.0004

ca-D' is D’ in cases, co-D' is D’ in controls, ca-r^2 is r^2 in cases and co-r^2 is r^2 in controls.

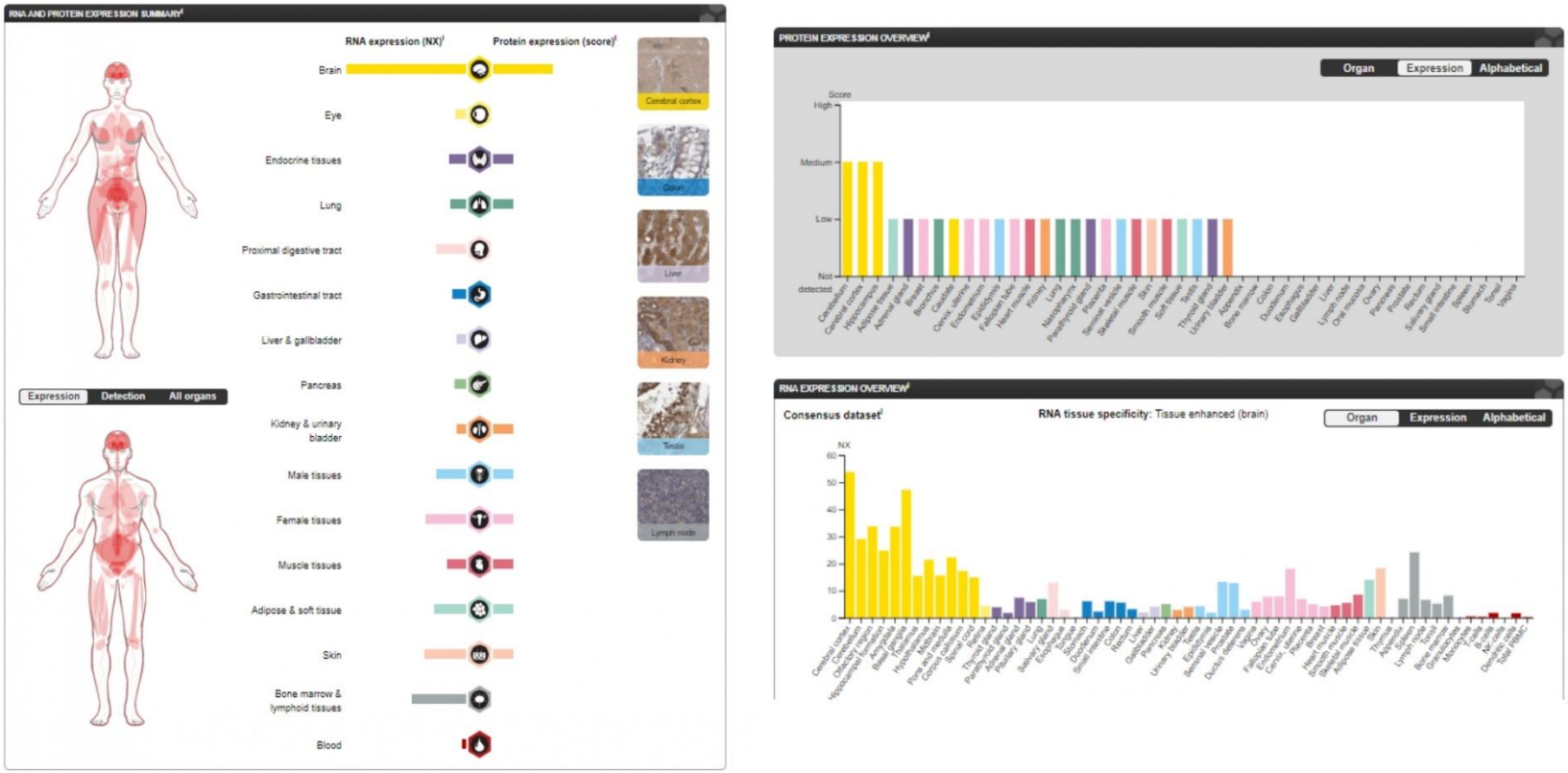
Supplementary Figures



Supplementary figure S1. Single marker association results within the major histocompatibility complex (MHC) in Chronic fatigue syndrome/Myalgic encephalomyelitis (ME/CFS) patients including imputed SNP data. The significance of the association signal, -log10P-value (y-axis), is plotted against the position on chromosome 6 shown in base pairs (bp, x-axis). The black, horizontal dashed line represents a significance threshold of P=0.001, while the grey dashed line represents the suggestive threshold of P=0.005. Association results are colored according to single nucleotide polymorphisms (SNPs), allelic HLA analyses, HLA amino acid analyses and HLA locus-wise analyses. Positions are according to National Center for Biotechnology Information's build 37 (hg19).



Supplementary figure S2. eQTL violin plots for significantly differentially expressed genes according to the rs4711249 genotypes in multiple tissues as reported in the GETx database.



Supplementary Figure S3. Overview of RNA and protein expression in various tissues for the GABBR1 gene taken from the human protein atlas

II

Genetic association study in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) identifies several potential risk loci

Riad Hajdarevic[1,2*], Asgeir Lande[1], Jesper Mehlsen[3], Anne Rydland[1,2], Daisy D. Sosa[4,5], Elin B. Strand[6,7] ,Olav Mella[8], Flemming Pociot[9], Øystein Fluge[8], Benedicte A. Lie[1,2,10,] Marte K. Viken[1,10*]

[1]Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway,

[2]Institute of Clinical Medicine, University of Oslo, Norway,

[3] Section for Surgical Pathophysiology, Rigshospitalet, University of Copenhagen, Denmark

[4]National Advisory Unit on CFS/ME, Norway,

[5]CFS/ME Center, Oslo University Hospital, Norway,

[6]Department of Digital Health Research, Division of Medicine, Oslo University Hospital, Oslo, Norway,

[7]Faculty of Health, VID-Specialized University, Diakonhjemmet Hospital, Oslo

[8]Department of Oncology and Medical physics, Haukeland University Hospital, Norway

[9]Department of Clinical Research, Steno Diabetes Center Copenhagen, Denmark,

[10]Department of Immunology, Oslo University Hospital and University of Oslo, Oslo, Norway

Key words: ME/CFS, Immunochip, GWAS

[*] Corresponding author: Riad Hajdarevic riad.hajdarevic@medisin.uio.no, Marte Viken m.k.viken@ous-research.no

Abstract

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is a disease of unknown etiology and pathogenesis, which manifests in a variety of symptoms like post-exertional malaise, brain fog, fatigue and pain. Hereditability is suggested by increased relatedness,

however, genome-wide association studies in ME/CFS have been limited by small sample sizes and broad diagnostic criteria, therefore no established risk loci exist to date. In this study, we have analyzed three ME/CFS cohorts: a Norwegian discovery cohort (N=427), a Danish replication cohort (N=460) and a replication dataset from the UK biobank (N=2105). To the best of our knowledge, this is the first ME/CFS genome-wide association study of this magnitude incorporating 2532 patients for the genome-wide analyses and 460 patients for a targeted analysis. Even so, we did not find any ME/CFS risk loci displaying genome-wide significance. Likewise, we did not replicate the findings from the Norwegian discovery cohort in either of our replication approaches. Nevertheless, several SNPs in the *TPPP* gene identified in the discovery cohort showed modest associations in the self-reported UK Biobank CFS cohort. Interestingly, the *TPPP* gene is expressed mostly in the brain and is associated with traits like Alzheimer's disease and multiple sclerosis. Considering the overall state of genetic studies in the ME/CFS field, we believe our well-designed approach represent an important step in the right direction in ME/CFS research.

INTRODUCTION

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is characterized by persistent, unexplained fatigue, post-exertional malaise as well as muscle pain and cognitive impairment. With a prevalence ranging from 0.2% to 2% (depending on diagnostic criteria), it represents a debilitating and serious medical condition affecting millions of individuals worldwide [1].

There is accumulating evidence pointing towards the involvement of the immune system in ME/CFS. Onset after viral exposure (Epstein-Barr-virus and human papillomavirus), as well as T cell alterations and autoantibodies have all been reported in ME/CFS [2] [3] [4, 5]. We have previously reported association with certain HLA class I and II variants [6] [7], which also is a hallmark of diseases where the immune system is involved. Furthermore, comorbidity of ME/CFS with various autoimmune diseases (AIDs) has been observed in a substantial number of patients [3]. In addition, the presence of hereditary components in ME/CFS is supported by excess relatedness[8] [9] [10] [11] pointing towards genetic risk factors being involved.

Several studies have reported genetic associations in ME/CFS, however, no consistent findings have been identified to date. Recently, associations with the *PTPN22* (rs2476601, P= 0.016) and *CTLA4* (rs3087243, P= 0.001) genes were reported in a candidate SNP study of 232 patients who developed ME/CFS after infection [12]. A candidate SNP study from 2006, of 43 patients, reported four SNPs in the *NR3C1* gene to reach significance ($p<0.05$) [13]. In addition, two independent genome wide association studies (GWAS), both comprising roughly 40 ME/CFS patients, reported several significant SNPs in immunologically relevant genes including the T cell receptor (TCR) alpha locus [14] [15]. Analysis of 353 ME/CFS patients who underwent genotyping by the company "23andme" also highlighted genes in immune pathways [16]. Nevertheless, no overlapping results were observed between any of the studies. These inconsistencies can be attributed to the lack of well-characterized phenotypes as well as the lack of large cohorts with statistical power [17].

Genome wide association studies (GWAS) have brought new insight into the genetics of many complex diseases, including autoimmune disorders (AIDs), revealing a genetic architecture characterized by hundreds of risk factors, generally with small effect sizes [18]

[19] [20] [21]. Furthermore, GWAS have demonstrated that many AIDs share a substantial amount of genetic risk factors with each other [22] [23], which lead to the development of the immunochip [24]. The establishment of genetic risk factors (i.e. $P<5x10^{-8}$) has been achieved by first conducting independent studies before enabling large consortium studies and meta-analyses of tens of thousands of patients and controls [25]. The genetic mapping of ME/CFS is thus lagging far behind, and few large patient cohorts have yet been established, leaving this research area in its infancy.

In an attempt to bring the ME/CFS genetics field forward, we performed a GWAS, based on the immunochip genotyping array covering AID risk loci, employing one ME/CFS discovery cohort and two independent replication cohorts incorporating in total more than 3000 ME/CFS patients.

MATERIAL AND METHODS

Study population

Our discovery cohort comprised 427 Norwegian ME/CFS patients collected from four separate inclusion groups; 1) the Rituximab study [26] [27], 2) the cyclophosphamide study [28], 3) the CFS/ME biobank at Oslo University Hospital and 4) patients recruited via announcements in patient networks. Duplicate patient samples in the different inclusion groups were removed, as well as patients with self-reported non-Norwegian/Scandinavian ancestry. All but four patients were diagnosed according to the 2003 Canadian Consensus Criteria (CCC) [11], these four fulfilled the 2010 International Consensus Criteria [29]. For the Norwegian patients, 14.3% (58/407) reported comorbidity with AIDs, with Hashimoto's thyreoditis/hypothyreosis and psoriasis being the most common. A total of 972 healthy, ethnically matched controls were also included in this study. The patients have been used previously in genetic studies of the major histocompatibility complex with less or different controls using actual HLA genotyping [6] [7].

In addition, we had two independent replication cohorts, from Denmark and the UK. The Danish cohort consisting of 460 ME/CFS cases diagnosed according to the CCC and 1965 anonymous Danish controls were available. From UK biobank, we extracted 2105 individuals registered with the field code 1182 (chronic fatigue syndrome) and 4786 gender matched, randomly selected controls from the UK biobank (genetic data release version 3 with imputation using the 1000 Genomes Project). To avoid population bias, only individuals who self-identify as "British," "Irish," "Any other white background" and "White," were included for analysis. Access to the UK biobank data was granted through UK biobank application 43949, and we have complied with all relevant ethical regulations for work with UK biobank, and all participants provided informed consent. The diagnostic criteria are self-reported in this cohort.

SNP genotyping

SNPs were genotyped using the Infinium ImmunoArray-24 v2 BeadChip (iChip v2, Illumina, San Diego, US) for the Norwegian ME/CFS patients at the Genomics Core Facility, Oslo

University hospital, while the healthy Norwegian controls were genotyped using HumanImmuno-v1 BeadChip (iChip v1), as described previously [30, 31]. The data was merged using Plink v1.9. Only the autosomal chromosomes were included in the analyses.

We aimed to replicate in our Danish cohort, 20 regions showing P<0.0003 in the discovery cohort. Two regions could not be included due to design issues. The remaining 18 candidate regions were covered by 24 tag SNPs, and genotyping was performed using Open Array Taqman technology on the QuantStudio 12K Flex Real-Time PCR System (Applied Biosystems). The genotypes for the Danish controls were obtained from iChip v1 data (Illumina), therefore we selected tag SNPs among the SNPs available on this array [32].

Quality control and statistical analysis

For all data sets, SNPs with genotyping success rate <99%, minor allele frequency <1%, and deviating from Hardy Weinberg equilibrium (p<0.001) in controls were excluded from analyses. Manual inspection of the iChip v2 genotyping cluster-plots for the Norwegian ME/CFS patients was performed, and poor performing SNPs were excluded. Only SNPs that were successfully genotyped and present on both iChip v1 and v2 were included. Thus, a grand total of 105,902 SNPs passed quality control and were included in the analyses of the Norwegian discovery cohort. The Michigan imputation server was used for SNP imputation (Reference Panel: 1000G Phase 3v5 EUR, rsq filter R<0.3, phasing via Eagle v2.4, Build 37) [33]. A principal-component analysis was performed using Plink v1.9 for the Norwegian and UK biobank samples and visualized using R (ggplot2) to ensure ethnically matched samples, no duplicates, and no close relatives in the data sets (Supplementary figure 1). Meta-analysis was done using Plink v1.9. For linkage disequilibrium (LD) plots, Haploview version 4.2 was used. We used a P-value threshold of <0.0003 in the discovery cohort to bring forward to replication. Thereafter, we report all P-values obtained from the different analyses, as the ME/CFS field is currently underpowered to reach genome-wide significance ($P<5x10^{-8}$) which is required to conclude that a locus is involved in predisposition to a given disease. Our data presented herein can be utilized in larger meta-analyses in order to reach this goal.

Databases used for gene expression

The GTEx database (v8) was used to investigate if the associated SNPs could serve as expression quantitative trait loci (eQTL). The data used for the analyses described in this manuscript were obtained from single tissue eQTL using the GTEx Portal on 30/07/21. The web tool Fuma was used to obtain gene expression data for specific genes in various tissues [34], as well as the DICE (Database of Immune Cell Expression, Expression quantitative trait loci (eQTLs) and Epigenomics) database [35]. In addition, we used Locus Focus for colocalization plots [36].

RESULTS

We first investigated the discovery cohort, where we had iChip array data from ME/CFS patients diagnosed according to the stringent Canadian consensus criteria (427 Norwegian ME/CFS cases and 972 Norwegian controls). After quality control, we included imputed SNP genotypes in the association analyses, thus increasing the dataset from 105,902 (Supplementary figure 2) to 1,462,996 SNPs. None of our associations reached genome-wide significance, however, 52 SNPs at chromosomes 5, 10, 12 and 13 were associated at a suggestive genome-wide significance level ($P<1x10^{-5}$) (Figure 1, Supplementary table 1). The most significant association signal was observed on chromosome 5, tagged by a directly genotyped (non-imputed) SNP, rs115523291, in the *TPPP* gene (2.5% in cases vs 0.4% in controls, $P= 8.5x10^{-7}$). The remaining regions displaying $P< 1x10^{-5}$ spanned *UBE2D1* (rs117354281, $P=8.4x10^{-6}$), *STAB2* (rs11111735, $P=1.6x10^{-6}$) and *LINC00333* (rs9546628, $P=3.6x10^{-6}$). In order to explore if associations were restricted to, or driven by, clinical sub-phenotypes, we performed association analyses on different subgroups of patients (i.e. autoimmune comorbidities, onset after infection or onset after vaccination), however, no consistent and significant differences were evident (Supplementary table 2 and 3). Using a threshold of $p<0.0003$, 18 regions were selected for replication, and we selected 24 SNPs that had been genotyped, and not imputed, as tag SNPs for these regions.

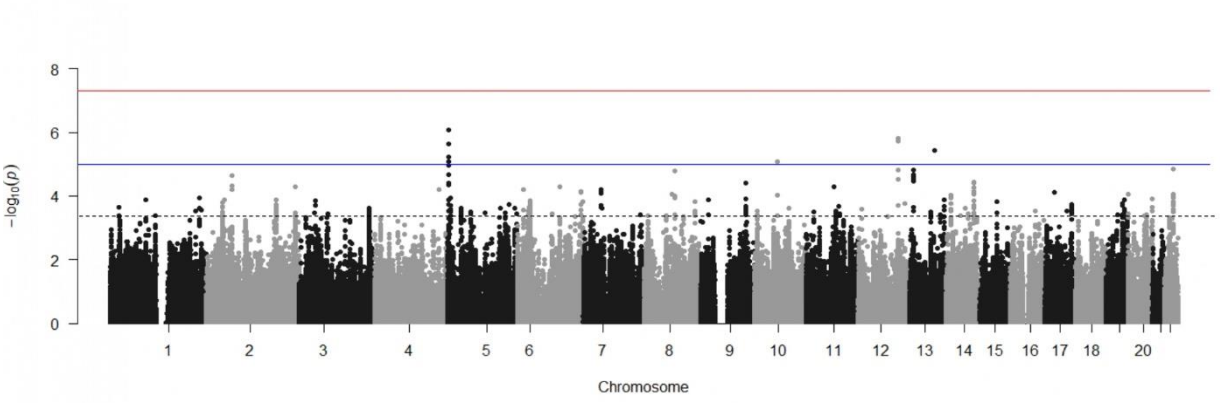


Figure 1. *SNP association results across the 22 autosomes in 427 Norwegian myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) patients and 972 healthy controls. The statistical significance of the association analysis as -log10 of the P-value (y-axis), is plotted against the chromosomal position of each chromosome in base pairs (bp, x-axis). The red, horizontal line represents a genome wide significance threshold of P=5 x10-8, while the blue line represents the suggestive significance level of P=1x10-5 and the dotted grey line represents the inclusion threshold for replication (P= 0.0003). Positions are according to National Center for Biotechnology Information's build 37 (hg19).*

No replication of tag SNPs

The selected 24 SNPs were first genotyped in the Danish ME/CFS patients. Two of the 24 SNPs failed genotyping and two did not pass quality control, leaving 20 SNPs (covering 15 suggestive regions) for replication analysis in the Danish cohort (460 cases and 1965 controls). Only one SNP, rs2453836, showed a p-value less than 0.05 without correcting for multiple testing (Supplementary table 4). Subsequently, we extracted genotypes for these 20 SNPs from the UK biobank (2105 self-reported CFS cases and 4786 controls), where four SNPs showed a p-value less than 0.05, namely rs2582085, rs115523291, rs8108136, rs6089982 (Supplementary table 4). Hence, none of the tag SNPs showed significant associations in the Danish or UK biobank cohorts. The odds ratios showed trends in the same direction in all three datasets for *ZBTB46* (rs6089982), *LINC00333* (rs7989859) and *IZUMO1* (rs8108136), (Figure 2), while the odds ratio deviated in the Danish ME/CFS cohort for *TPPP* (rs115523291). However, this might be due to chance as the confidence intervals tend to cross

OR=1 for one or both of the replication cohorts. For the remaining SNPs, the results showed much larger differences between the datasets (Supplementary figure 3).

In the combined analysis of the datasets, all SNPs displayed, in general, less significant P-values (Supplementary table 5) than seen in the Norwegian discovery cohort. Notably, none of the SNPs reached the genome-wide significance or suggestive threshold. Only 12 of the 20 SNPs showed association at a $P \leq 0.05$, without correcting for multiple testing. The most significant associations were seen for rs6089982 in *ZBTB46* (P=0.0003), rs7989859 in *LINC00333* (P=0.0005), rs115523291 in *TPPP* (P=0.002) and rs8108136 in *IZUMO1* (P=0.002)

Since the diagnosis in the UK biobank cohort, in contrast to the Norwegian and Danish cohorts, was self-reported and not diagnosed using the Canadian consensus criteria, we also performed association analyses after filtering out illnesses (autoimmune and psychiatric diseases) that could potentially confound the diagnosis and reanalyzed the UK dataset. However, this did not alter the initial results (Supplementary table 6).

No replication of multiple SNPs across investigated regions

A weakness of our initial replication approach was that it relied on the ability of the selected tag SNPs to capture an association within each region across all datasets. Since the Norwegian and UK cohorts had genome-wide data, this enabled us to use an alternative replication strategy where we examined all SNPs across the regions implicated by the discovery analysis. In accordance with the LD structure in the regions selected for replication, the association plots for the Norwegian cohort showed that several SNPs supported the ME/CFS associations (Supplementary figure 4).When including the imputed SNPs, for some regions markedly stronger associations were seen in the discovery cohort with imputed SNPs than the tag SNPs we had initially selected for replication, particularly for chromosome 5 (*CEP72, TPPP*), chromosome 14 (*RIN3*) and chromosome 22 (*CACNA11*). Furthermore, in the UK dataset, the tag SNPs represented the regional association signal even worse, and for most regions the tag SNP was far from being the most associated SNP (Supplementary figure 5). Nevertheless, for eight of the 15 inspected regions, other SNPs (<400 kb away from our tag SNPs) showed

ME/CFS association with P values <0.001. Therefore, we next combined the Norwegian and UK dataset to investigate all available SNPs across all regions implicated by the discovery analysis (Supplementary figure 6). The regions showing the strongest association in the combined dataset (Figure 3), with their respective novel top SNPs, were: *TPPP* (rs139264145; P= 0.00004), *LINC00333* (rs368711309; P= 0.002), *RIN3* (rs4904960; P= 0.0003), *IGFBP1/IGFBP3* (rs28552707; P= 0.002), *IZUMO1* (rs28745910; P= 0.0002) and *ZBTB46* (rs2777943; P= 0.0002). Hence, these regional replication analyses pointed out two potentially novel loci, in addition to those revealed in the single tag SNP analysis, namely *RIN3 and IGFBP1/IGFBP3.*

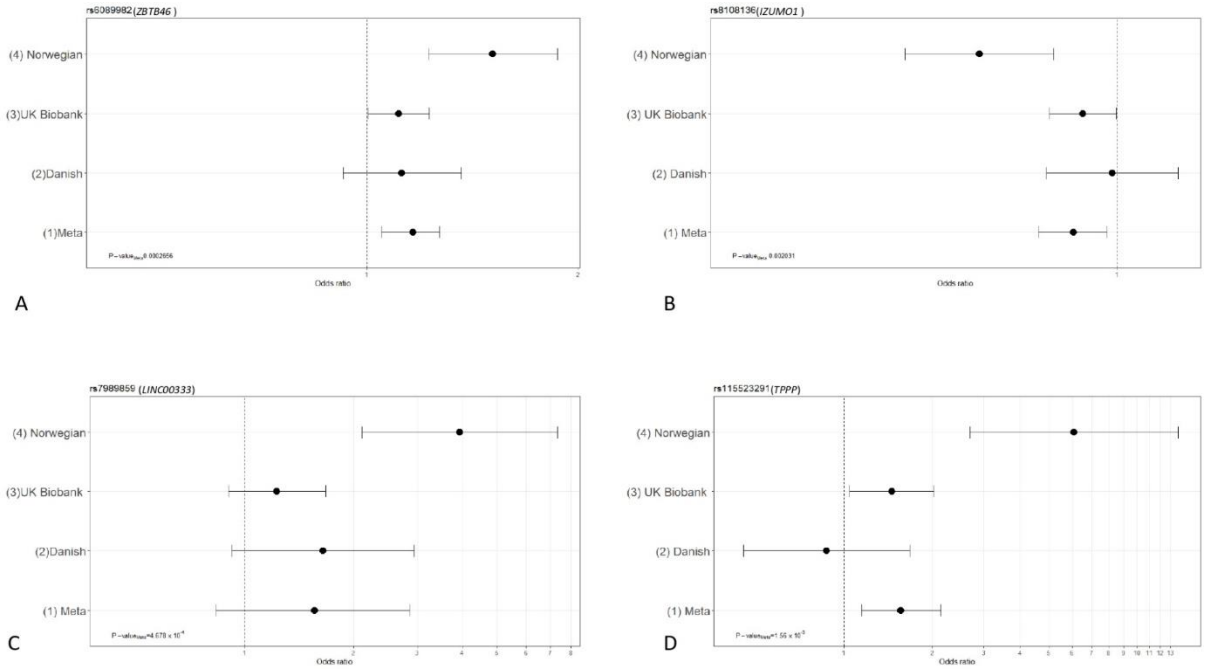


Figure 2. *Odds ratios plots for the tag SNPs rs6089982, rs8108136, rs7989859 and rs115523291 with the odds ratio plotted separately for the Norwegian cohort (N=427 cases and N=972 controls), UK biobank cohort (N=2105 cases and N=4786 controls) and the Danish cohort (N=460 cases and N=1956 controls) and from the meta-analysis of all three cohorts.*

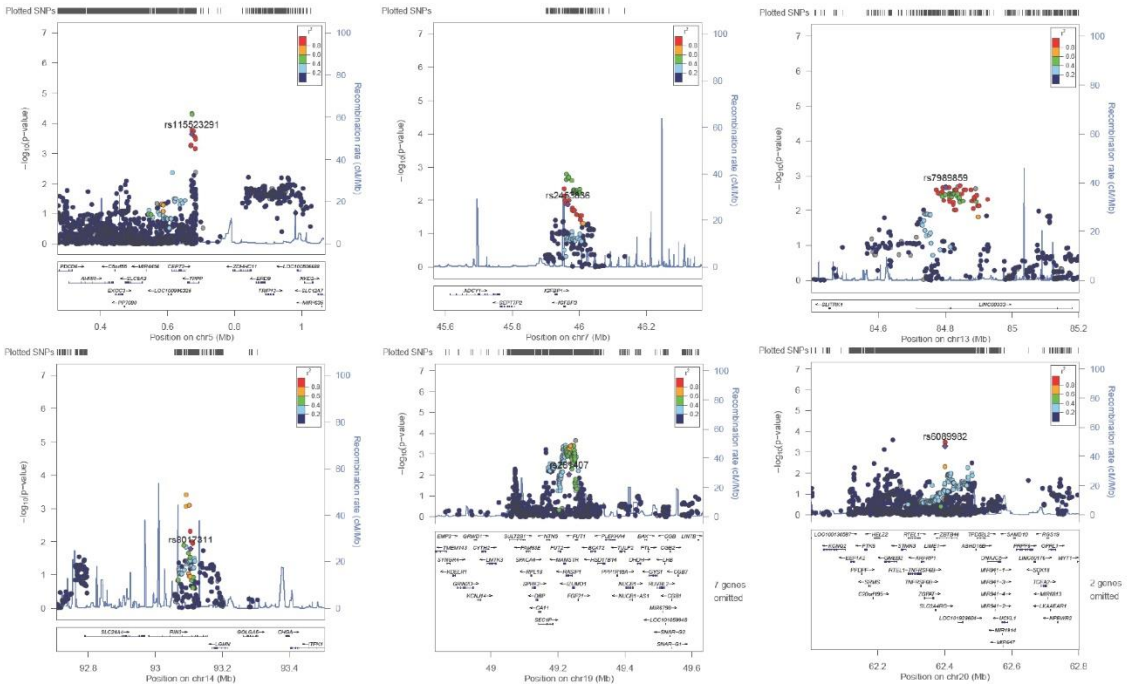


Figure 3. *Locus zoom plots for the regional analyses of e the putative ME/CFS associated regions identified in the Norwegian discovery cohort (TPPP, LINC00333, RIN3, IGFBP1/IGFBP3, IZUMO1 and ZBTB46). The plots show the meta-analysis results of the Norwegian cohort (N=427 cases and N=972 controls) and UK biobank cohort (N=2105 cases and N=4786 controls) including imputed SNPs.*

Association signals derived from the large UK biobank dataset

Due to the superior power of UK biobank, we next used the UK dataset as the discovery cohort (Supplementary table 7). This revealed 228 SNPs with suggestive significance ($P<1x10^{-5}$) and one SNP upstream of *NOX3* (NADPH oxidase 3) on 6q15 reached GWAS level significance (rs77381650, $p= 4.4x10^{-8}$). We then extracted data for these 228 SNPs from our Norwegian cohort, but only 10 were present among either the genotyped or imputed SNPs. However, none of the ten overlapping SNPs showed any signs of association (P>0.1).

In light of these observations, we investigated if our initial approach had overlooked other regions by performing a combined analysis of all SNPs across the genome available from the Norwegian and UK Biobank cohorts. This analysis revealed six novel associated regions ($P<1x10^{-5}$), three of these with several associated SNPs spanning *EPHA7, SKAP1* and *SHANK3* (Supplementary table 8). The P-values for these SNPs were all <0.0001 in the UK

biobank, while most were non-significant in the Norwegian cohort, where the strongest P-values were seen for SNPs in *SKAP1* (P=0.02).

The possible implication of the *TPPP* region in ME/CFS

Several SNPs encompassing the *TPPP* gene showed association both in the Norwegian (Figure 4a) and the UK (Figure 4b) dataset. In the Norwegian cohort, the rare SNP rs115523291 displayed the peak association signal, while the more common SNP rs451979 was the most significantly associated in the UK dataset (57.2% in cases vs 54.2% in controls, P=0.001). However, the infrequent and associated tag SNP from the Norwegian dataset, rs115523291, also showed an association in the UK dataset (1.4% in cases vs 1.0% in controls, P=0.03) and in the combined dataset (P=0.0002). Overall, the most significant association in the combined analyses of the Norwegian and UK datasets was with rs139264145 (OR=1.9; $P=4.7x10^{-5}$), a SNP in strong LD with rs115523291 (Figure 4c) in both the Norwegian ($r^2=1$) and UK dataset ($r^2=0.88$). These two SNPs were equally associated in the Norwegian ME/CFS discovery cohort ($P= 8.46x10^{-7}$) and likely represent the same association signal (Supplementary table 1). There was no correlation ($r^2<0.01$) between the top SNP (rs139264145) and the most significant SNP from the UK biobank (rs451979), but strong LD measured by D'=1, indicating haplotype structures (Figure 4d). Haplotype analyses of rs451979 and rs139264145 showed global associations in both the Norwegian ($P= 9.4x10^{-6}$) and the UK dataset (P=0.0005), but with different individual haplotypes driving the association.

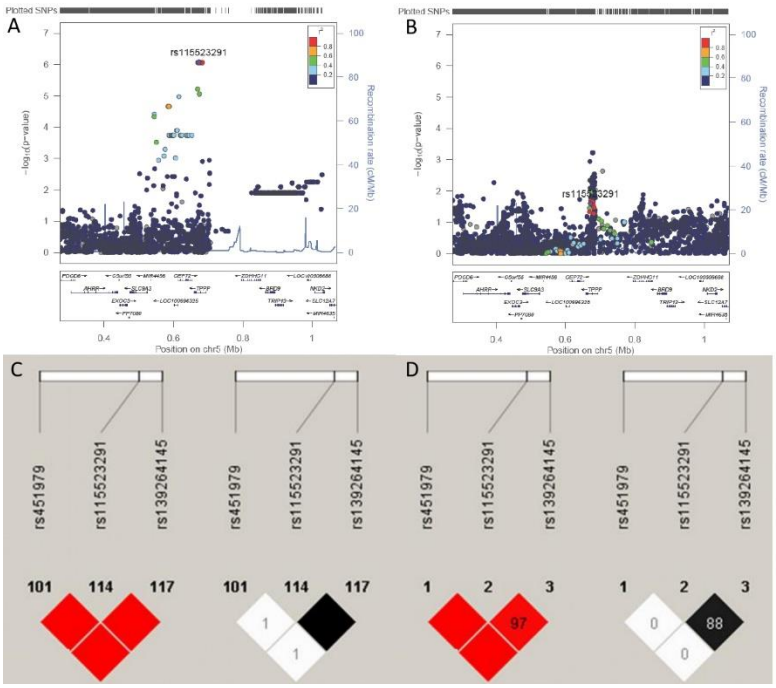


Figure 4. *Locus zoom plots of the tag SNP rs115523291 in (A) the Norwegian cohort (N=427 cases and N=972 controls) and (B) in the UK Biobank cohort (N=2105 cases and N=4786 controls). Linkage disequilibrium (LD) plot for SNPs rs451979, rs139264145 and rs115523291 measured by r^2 in the Norwegian cohort(C) and the UK biobank cohort (D). Values shown in the figures are 100 x r^2 and 100 x D' (C and D).*

No replication of previously reported associations with the *PTPN22 and CTLA4* genes

We also investigated the recently reported associations with the immunologically relevant *PTPN22* (rs2476601) and *CTLA4* (rs3087243) genes in patients who developed ME/CFS triggered by infection [12]. We found no associations in the Norwegian dataset (P= 0.9 for rs2476601 and P=0.2 for rs3087243) even after only including patients who reported disease onset after infection (P>0.6; Supplementary table 9). No association was observed with overall CFS in the UK biobank data either (P=0.7 for both rs2476601 and rs3087243; Supplementary table 9). A combined analysis of all SNPs across these genes in our Norwegian and UK dataset (Supplementary figure 7) showed some association signals with other SNPs in both the *PTPN22* (P≥0.004) and *CTLA4* (P≥0.0007) genetic regions.

Annotation of the putative ME/CFS regions

Finally, we aimed to annotate functional relevance on our top and tag SNPs in the regions unraveled as potentially ME/CFS associated. Colocalization analyses showed potential

colocalization for SNPs in the *IZUMO1* region with the eQTL signals from relevant tissues like brain and lung (Figure 5, Simple sum $p=8.4x10^{-4}$) in the joint analysis of the Norwegian and UK biobank cohorts (See supplementary figure 8 for colocalization plots for all candidates).

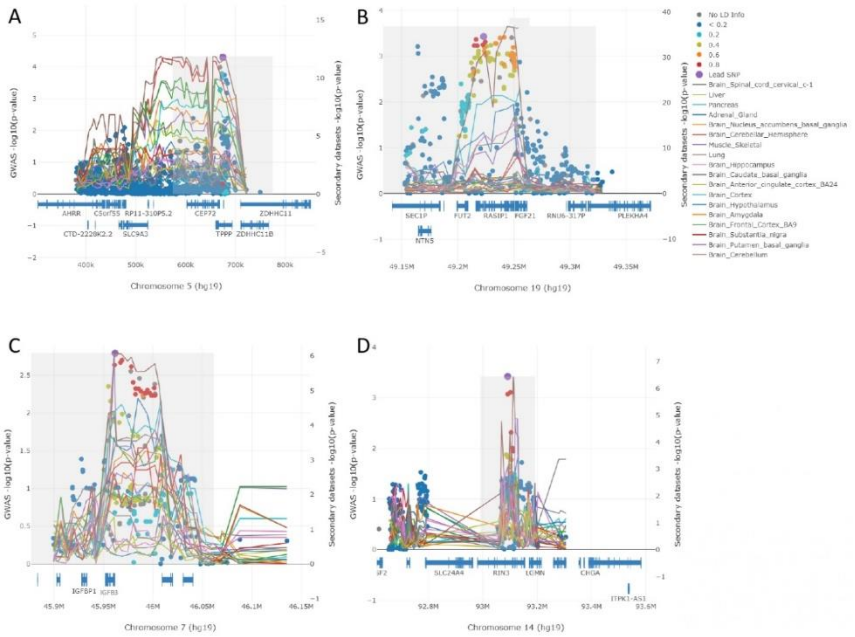


Figure 5. *Colocalization plot generated in LocusFocus for the combined analysis of the Norwegian cohort (N=427 cases and N=972 controls) and the UK biobank cohort (N=2105 cases and N=4786 controls) showing the degree of pairwise LD with the lead markers (rs139264145, rs28552707, rs28745910). The x-axis represents the chromosomal position in base pairs while the two y-axes represent the significance of the SNP and eQTL associations respectively*

When exploring the GTEx database, the tag SNP rs8108136 showed significant eQTL results in the investigated tissues (blood, brain, skeletal muscle and nerve) with correlations to the gene expression of *IZUMO1, MAMSTR, NTN5* and *FUT2* (Figure 5A and supplementary figure 9). In whole blood, the tag SNP rs2453836 genotype influenced the gene expression level of *IGFBP3* (Figure 5B), while the top SNP in the same region showed association with expression of both *TTC4P1* and *FTLP15* in skeletal muscle tissue (Supplementary figure 8). The last SNP that had significant eQTL results was the top SNP in the *ZBTB46* region, rs2777943, which was associated with expression levels of *STMN3* in blood (Figure 5B). Additionally, SNPs in the *SKAP1* gene, identified in the whole genome meta-analysis of the

Norwegian and UK Biobank datasets, were implicated as eQTLs with several neighboring genes expressed in relevant tissue like muscle and lung (Supplementary figure 10).

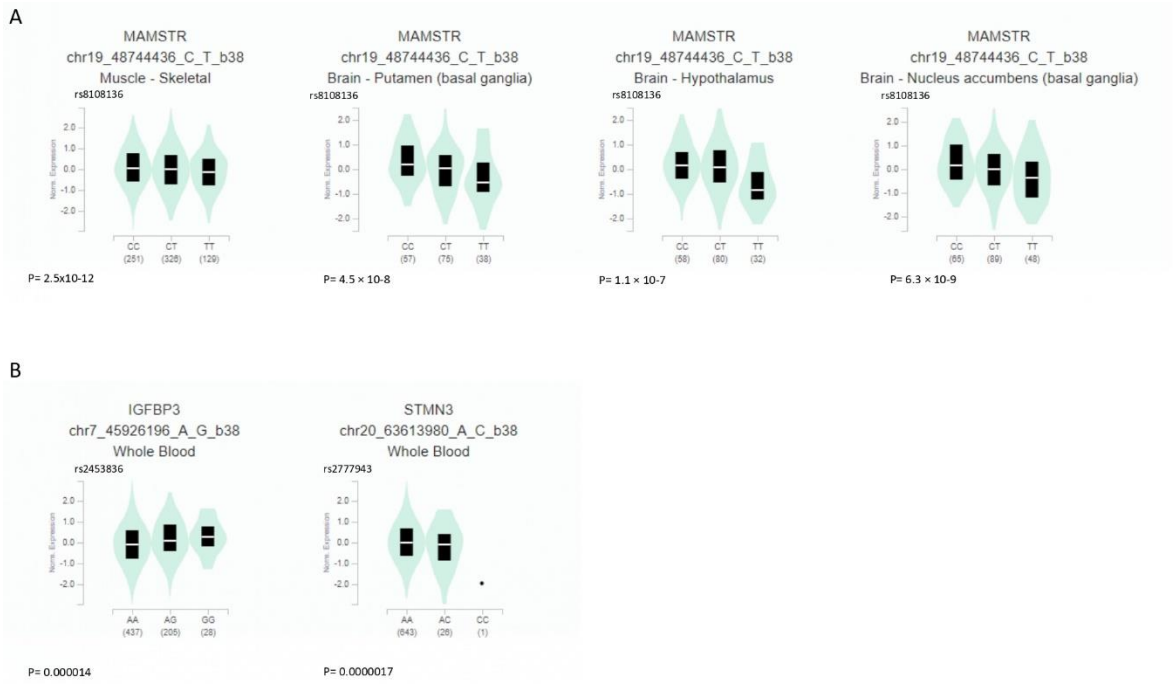


Figure 6. *Expression quantitative trait loci (eQTL) plots between genes showing high expression in the relevant tissues (obtained from GTEx v8) and A) for tag SNP rs8108136 and B) tag SNP rs2453836 and rs2777943.*

Several of the annotated genes are expressed in brain tissues (Figure 6A), with *TPPP, ZBTB46, STMN3 and MAMSTR* showing high expression levels. Few genes appeared to be expressed in whole blood, only *RIN3* and *STMN3* showed pronounced expression (Figure 6A). Furthermore, immune cells, including monocytes, B cells and several T cell subsets, express our candidate genes (Supplementary figure 11), in particular *RIN3, STMN3* and *IGFBP3* (Figure 6B).

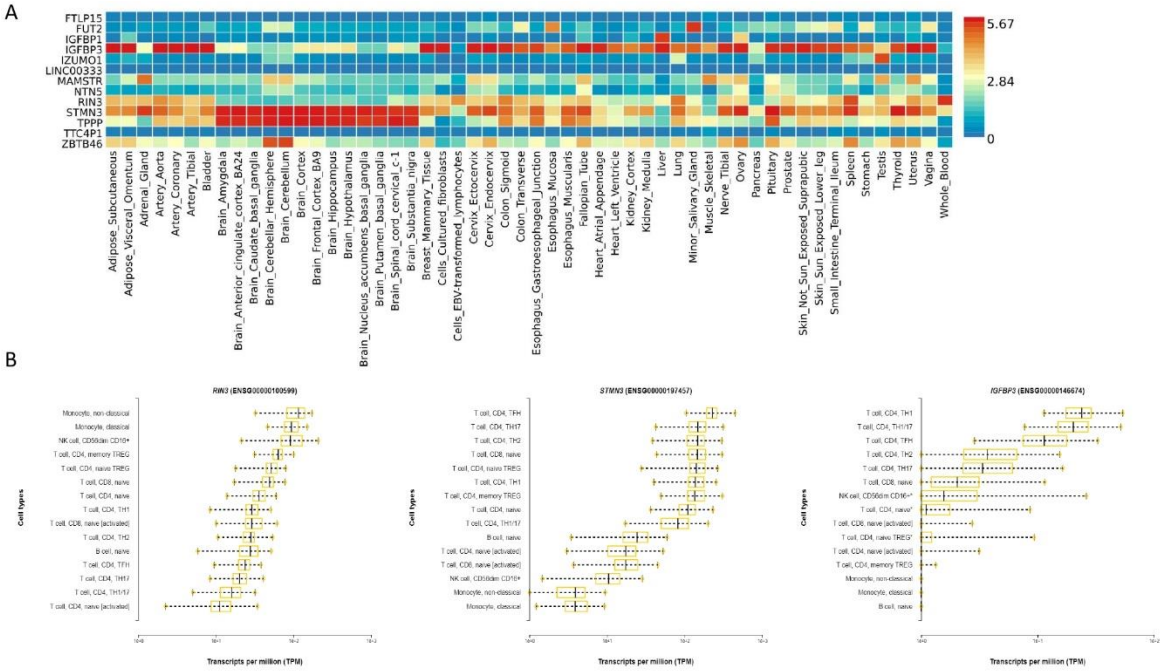


Figure 7. A) *Gene expression levels in various tissues of candidate genes selected either by their proximity to tag and top SNPs, or due to their involvement in eQTLs (Obtained via Fuma). B) Expression levels in immune cells of the genes most highly expressed in blood*

DISCUSSION

Using genome-wide array data and large ME/CFS cohorts (>2900 patients in total), we have identified several chromosomal regions with suggestive ME/CFS associations that warrants follow-up in subsequent studies towards the future establishment of the first ME/CFS genetic risk loci at genome-wide significance.

We used different replication approaches due to restrictions imposed by the available DNA and datasets. In the first strategy, we selected tag SNPs from our Norwegian discovery analysis to be replicated in both the Danish cohort and in the UK dataset. Notably, we only selected directly genotyped SNPs as tag SNPs (in order to match the genotypes in the Danish control dataset), which could partly explain the observed discrepancies between the cohorts as these tags might not be in sufficient LD with the putative ME/CFS risk SNPs in the different populations. We aimed to overcome this by using regional analyses of all available SNPs across the regions to be replicated. The draw-back of the latter approach was that we could not include the Danish cohort. Both approaches, nevertheless, pointed out the same four loci (*TPPP, LINC00333, IZUMO1* and *ZBTB46*), but the latter strategy also picked up signals from two additional loci (*RIN3* and *IGFBP1/IGFBP3*) to be possibly followed up. Nevertheless, it must be stated that given our lack of statistical power those observations could be false positive.

In our second replication approach, we used the UK Biobank cohort as a discovery cohort and the Norwegian as replication. However, this approach did not reveal any overlap between the two datasets since only ten SNPs of the 228 most associated UK Biobank SNPs were present in the Norwegian cohort. Additionally, we investigated available whole genome data from the Norwegian and UK Biobank cohorts combined. However, due to the substantial size differences of the two cohorts and divergent phenotyping, this introduced biases towards associations driven by the UK Biobank cohort.

Our most significant finding was with SNPs encompassing the *TPPP* region. *TPPP* SNPs showed associations in both the Norwegian and the UK dataset. However, the strongest association signals were seen with different SNPs, i.e. a rare variant in the Norwegian cohort and a more common variant in the UK dataset. However, the LD analyses indicated clear haplotype patterns, and globally haplotype analyses showed association in both cohorts. This

could potentially indicate that these SNPs pick up a common causal risk variant that had not been included in our current analyses. Notably, the region just centromeric of the association peak was poorly covered by SNPs in our combined meta-analysis. Interestingly, the *TPPP* gene, encoding the tubulin polymerization promoting protein, is mainly expressed in the brain. The TPPP protein plays a pivotal role in the myelination of oligodendrocytes [37] and has been shown to correlate with shortened disease duration in multiple sclerosis [38]. This may indicate a role for TPPP in myelin repair. Hence, changes in this gene may underlie neurological abnormalities and may be involved in pathologies like Alzheimer's disease. [39]. Furthermore, *RIN3* upregulation has recently been reported in Alzheimer mouse models [40]. In addition, *IGFBP1* and *FUT2* have been associated with Crohn's disease and rheumatoid arthritis, respectively [41] [42]. However, most intriguingly the downregulation of the *MAMSTR* gene has been reported to correlate with impaired skeletal muscle regeneration in mice [43], which is interesting given the ME/CFS clinical manifestation of muscle pain and fatigue. Several of the annotated genes in our putative ME/CFS associated regions have expression patterns in cells and tissues that could be compatible with a functional role in ME/CFS.

We did not replicate previously reported associations with the immunologically important genes, *PTPN22* and *CTLA4* [12]. These genes have been found to predispose for a large number of autoimmune diseases. We found no evidence of association with the presumed causal SNPs in our dataset, but some evidence of association with surrounding SNPs not being in LD. Notably, these SNP associations have previously been reported to be restricted to infection-triggered ME/CFS, and history of bacterial or viral infection was collected from patient records.[12] We only had self-reported information about infection episodes prior to disease onset from our Norwegian cohort, and not the entire dataset, but found no evidence of association in this patient strata. Our self-reported infection-trigger is likely less reliable, however, ideally both studies should have had serological confirmation. The putative involvement of these autoimmune loci in certain ME/CFS sub-phenotypes should be further addressed in future studies.

Due to the clinical heterogeneity of ME/CFS as well as a presumed multifactorial aetiology, genetic risk variants can be assumed to have a small effect size. The associations we observed

had odds ratios in line with this notion (OR < 1.6). To obtain the desired power of 80%, we would need up to 10 times more patients [44]. This is, however, a conservative estimate which does not consider the obvious heterogenic nature of ME/CFS. In this study, we incorporated a total of 887 patients diagnosed via the Canadian Consensus Criteria and 2105 self-reported CFS cases. This is by far the largest ME/CFS study to date considering that similar studies had around 40 patients each [14] [15].

However, lessons from genome-wide associations in autoimmune and other complex diseases have taught us that several thousand patients and controls are needed to firmly reveal the risk loci [22] [25]. It is important to stress that we cannot exclude false positive or negative findings in our cohort given our statistical limitation and low allele frequencies of our top hits [45]. As has been the history of unraveling the genetic architecture of most common diseases, more GWASs are needed in order to internationally reach the number of patients necessary to provide sufficient power for meta-analyses to establish genetic associations with genome-wide significance.

Despite the current lack of large cohorts of phenotypically stringent and well-characterized ME/CFS patients, this study represents, to the best of our knowledge, the largest and most homogenous genetic study performed in ME/CFS so far. Ongoing projects like the DecodeME project (www.decodeme.org.uk) will enable future studies of larger and more powerful cohorts, which are warranted to produce the desired statistical power to definitively investigate the genetic architecture of ME/CFS.

In conclusion, we identified several potential risk loci for ME/CFS, which encourage further investigations. Future genetic studies should be performed in large cohorts of several thousand patients and strive to use strict and comprehensive phenotyping to enable analyses of homogenous sub-phenotypes.

Acknowledgements

We are very grateful to the patients who participated in the study. Thank you to Wenche Kristiansen and Hilde Haukeland at the CFS/ME Biobank, Oslo University Hospital, Hanne S. Hjorthaug and Siri T. Flåm at Department of Medical Genetics, Oslo University Hospital. We also thank the employees at the Genomics Core Facility, Oslo University hospital, Norway for all their help. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

Funding

Funding was provided by the Research Council of Norway (application number: 272681) and the Kavli trust. The content is solely the responsibility of the authors and does not necessarily reflect the views of the funders. The funders did not influence the research or the manuscript in any way.

Conflict of interest disclosure

The authors do not declare any conflict of interests.

1. Lim, E.-J., et al., *Systematic review and meta-analysis of the prevalence of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME).* Journal of Translational Medicine, 2020. **18**(1): p. 100.
2. Sepúlveda, N., et al., *Myalgic Encephalomyelitis/Chronic Fatigue Syndrome as a Hyper-Regulated Immune System Driven by an Interplay Between Regulatory T Cells and Chronic Human Herpesvirus Infections.* Frontiers in immunology, 2019. **10**: p. 2684-2684.
3. Sotzny, F., et al., *Myalgic Encephalomyelitis/Chronic Fatigue Syndrome – Evidence for an autoimmune disease.* Autoimmunity Reviews, 2018. **17**(6): p. 601-609.
4. Rasa, S., et al., *Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS).* Journal of translational medicine, 2018. **16**(1): p. 268-268.
5. Feiring, B., et al., *HPV vaccination and risk of chronic fatigue syndrome/myalgic encephalomyelitis: A nationwide register-based study from Norway.* Vaccine, 2017. **35**(33): p. 4203-4212.
6. Lande, A., et al., *Human Leukocyte Antigen alleles associated with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS).* Scientific Reports, 2020. **10**(1): p. 5267.
7. Hajdarevic, R., et al., *Fine mapping of the major histocompatibility complex (MHC) in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) suggests involvement of both HLA class I and class II loci.* Brain, Behavior, and Immunity, 2021.
8. Albright, F., et al., *Evidence for a heritable predisposition to Chronic Fatigue Syndrome.* BMC neurology, 2011. **11**: p. 62-62.
9. Lawrie, S.M. and A.J. Pelosi, *Chronic Fatigue Syndrome in the Community Prevalence and Associations.* British Journal of Psychiatry, 1995. **166**(6): p. 793-797.
10. Buchwald, D., et al., *A Twin Study of Chronic Fatigue.* Psychosomatic Medicine, 2001. **63**(6).
11. Carruthers, B.M., et al., *Myalgic Encephalomyelitis/Chronic Fatigue Syndrome.* Journal of Chronic Fatigue Syndrome, 2003. **11**(1): p. 7-115.
12. Steiner, S., et al., *Autoimmunity-Related Risk Variants in PTPN22 and CTLA4 Are Associated With ME/CFS With Infectious Onset.* Frontiers in Immunology, 2020. **11**(578).
13. Goertzel, B.N., et al., *Combinations of single nucleotide polymorphisms in neuroendocrine effector and receptor genes predict chronic fatigue syndrome.* Pharmacogenomics, 2006. **7**(3): p. 475-483.
14. Schlauch, K.A., et al., *Genome-wide association analysis identifies genetic variations in subjects with myalgic encephalomyelitis/chronic fatigue syndrome.* Translational psychiatry, 2016. **6**(2): p. e730-e730.
15. Smith, A.K., et al., *Convergent genomic studies identify association of GRIK2 and NPAS2 with chronic fatigue syndrome.* Neuropsychobiology, 2011. **64**(4): p. 183-94.
16. Perez, M., et al., *Genetic Predisposition for Immune System, Hormone, and Metabolic Dysfunction in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: A Pilot Study.* Frontiers in Pediatrics, 2019. **7**(206).
17. Dibble, J.J., S.J. McGrath, and C.P. Ponting, *Genetic risk factors of ME/CFS: a critical review.* Human Molecular Genetics, 2020. **29**(R1): p. R117-R124.
18. Parkes, M., et al., *Genetic insights into common pathways and complex relationships among immune-mediated diseases.* Nat Rev Genet, 2013. **14**(9): p. 661-73.
19. Matzaraki, V., et al., *The MHC locus and genetic susceptibility to autoimmune and infectious diseases.* Genome biology, 2017. **18**(1): p. 76-76.
20. Dendrou, C.A., et al., *HLA variation and disease.* Nature Reviews Immunology, 2018. **18**: p. 325.
21. Wang, L., F.-S. Wang, and M.E. Gershwin, *Human autoimmune diseases: a comprehensive update.* Journal of Internal Medicine, 2015. **278**(4): p. 369-395.
22. Kochi, Y., *Genetics of autoimmune diseases: perspectives from genome-wide association studies.* International immunology, 2016. **28**(4): p. 155-161.
23. Ricaño-Ponce, I. and C. Wijmenga, *Mapping of immune-mediated disease genes.* Annu Rev Genomics Hum Genet, 2013. **14**: p. 325-53.
24. Cortes, A. and M.A. Brown, *Promise and pitfalls of the Immunochip.* Arthritis Res Ther, 2011. **13**(1): p. 101.

25. Visscher, P.M., et al., *10 Years of GWAS Discovery: Biology, Function, and Translation.* Am J Hum Genet, 2017. **101**(1): p. 5-22.
26. Fluge, Ø., et al., *Benefit from B-lymphocyte depletion using the anti-CD20 antibody rituximab in chronic fatigue syndrome. A double-blind and placebo-controlled study.* PLoS One, 2011. **6**(10): p. e26358.
27. Fluge, Ø., et al., *B-Lymphocyte Depletion in Patients With Myalgic Encephalomyelitis/Chronic Fatigue Syndrome.* Annals of Internal Medicine, 2019. **170**(9): p. 585-593.
28. Rekeland, I.G., et al., *Intravenous Cyclophosphamide in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. An Open-Label Phase II Study.* Frontiers in medicine, 2020. **7**: p. 162-162.
29. Carruthers, B.M., et al., *Myalgic encephalomyelitis: International Consensus Criteria.* Journal of internal medicine, 2011. **270**(4): p. 327-338.
30. Beecham, A.H., et al., *Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis.* Nature Genetics, 2013. **45**(11): p. 1353-1360.
31. Liu, J.Z., et al., *Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis.* Nature Genetics, 2013. **45**(6): p. 670-675.
32. Barrett, J.C., et al., *Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes.* Nat Genet, 2009. **41**(6): p. 703-7.
33. Das, S., et al., *Next-generation genotype imputation service and methods.* Nature genetics, 2016. **48**(10): p. 1284-1287.
34. Watanabe, K., et al., *Genetic mapping of cell type specificity for complex traits.* Nature Communications, 2019. **10**(1): p. 3222.
35. Schmiedel, B.J., et al., *Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression.* Cell, 2018. **175**(6): p. 1701-1715.e16.
36. Panjwani, N., et al., *LocusFocus: Web-based colocalization for the annotation and functional follow-up of GWAS.* PLOS Computational Biology, 2020. **16**(10): p. e1008336.
37. Fu, M.M., et al., *The Golgi Outpost Protein TPPP Nucleates Microtubules and Is Critical for Myelination.* Cell, 2019. **179**(1): p. 132-146.e14.
38. Höftberger, R., et al., *Tubulin polymerization promoting protein (TPPP/p25) as a marker for oligodendroglial changes in multiple sclerosis.* Glia, 2010. **58**(15): p. 1847-57.
39. Frykman, S., et al., *Identification of two novel synaptic γ-secretase associated proteins that affect amyloid β-peptide levels without altering Notch processing.* Neurochem Int, 2012. **61**(1): p. 108-18.
40. Shen, R., et al., *Upregulation of RIN3 induces endosomal dysfunction in Alzheimer's disease.* Transl Neurodegener, 2020. **9**(1): p. 26.
41. Ye, B.D., et al., *Association of FUT2 and ABO with Crohn's disease in Koreans.* J Gastroenterol Hepatol, 2020. **35**(1): p. 104-109.
42. Li, J., et al., *TXNDC5 contributes to rheumatoid arthritis by down-regulating IGFBP1 expression.* Clinical & Experimental Immunology, 2018. **192**(1): p. 82-94.
43. Yanay, N., et al., *Pax7, Pax3 and Mamstr genes are involved in skeletal muscle impaired regeneration of dy2J/dy2J mouse model of Lama2-CMD.* Hum Mol Genet, 2019. **28**(20): p. 3369-3390.
44. Altshuler, D., M.J. Daly, and E.S. Lander, *Genetic mapping in human disease.* Science, 2008. **322**(5903): p. 881-8.
45. Shen, X. and O. Carlborg, *Beware of risk for increased false positive rates in genome-wide association studies for phenotypic variability.* Frontiers in genetics, 2013. **4**: p. 93-93.

Supplementary tables

Supplementary table1. Supplementary table 1: SNPs reaching suggestive significance ($P<1x10^{-5}$) in the Norwegian discovery cohort with statistics for the UK Biobank cohort presented for overlapping SNPs.

CHR	SNP	BP	AFCANOR	AFCONOR	PNOR	OR(95%CI)NOR	FCAUK	FCOUK	PUK	OR(95%CI)UK
5	rs148123612	674413	0.029	0.006	8.4X10-07	4.85(2.48-9.71)	NA	NA	NA	NA
5	rs146055587	670104	0.025	0.004	8.5X10-07	6.10(2.69-13.83)	0.01	0.009	0.04	1.40(1.00-1.95)
5	rs115523291	672911	0.025	0.004	8.5X10-07	6.10(2.69-13.83)	0.01	0.009	0.03	1.45(1.04-2.03)
5	rs116836284	673028	0.025	0.004	8.5X10-07	6.10(2.69-13.83)	0.01	0.009	0.02	1.48(1.05-2.07)
5	5:673273:G:GCAACT	673273	0.025	0.004	8.5X10-07	6.10(2.69-13.83)	NA	NA	NA	NA
5	rs139264145	673278	0.025	0.004	8.5X10-07	6.10(2.69-13.83)	0.01	0.009	0.009	1.57(1.11-2.21)
5	5:675578:CAG:C	675578	0.025	0.004	8.5X10-07	6.10(2.69-13.83)	NA	NA	NA	NA
5	rs116271924	677781	0.025	0.004	8.5X10-07	6.10(2.69-13.83)	0.01	0.009	0.02	1.47(1.05-2.05)
5	rs112297579	681490	0.025	0.004	8.5X10-07	6.10(2.69-13.83)	0.01	0.009	0.03	1.44(1.03-2.02)
5	rs149851820	683032	0.025	0.004	8.5X10-07	6.10(2.69-13.83)	0.01	0.01	0.05	1.38(0.99-1.91)
5	rs114126299	683822	0.025	0.004	8.5X10-07	6.10(2.69-13.83)	0.01	0.009	0.03	1.43(1.02-1.99)
12	rs11111735	1,04E+08	0.060	0.024	1.6X10-06	2.62(1.74-3.93)	0.1	0.12	0.1	0.90(0.80-1.02)
12	rs12424999	1,04E+08	0.059	0.023	1.9X10-06	2.62(1.74-3.93)	0.1	0.12	0.1	0.90(0.80-1.02)
5	5:641653_AAAC_A	641653	0.027	0.006	2.3X10-06	4.86(2.36-10.02)	NA	NA	NA	NA
13	13:84758808:T:C	84758808	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	NA	NA	NA	NA
13	rs9546628	84762873	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.05	0.05	0.82	1.01(0.86-1.19)
13	rs17077847	84772552	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.21	1.21(0.89-1.65)
13	rs76732893	84778444	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.02	0.01	0.22	1.20(0.88-1.63)
13	rs7324586	84792045	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.23	1.20(0.88-1.64)
13	rs114039149	84793091	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.19	1.22(0.90-1.67)
13	rs9593884	84795804	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.02	0.01	0.28	1.17(0.87-1.58)
13	13:84796408:A:G	84796408	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	NA	NA	NA	NA
13	rs7989859	84801379	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.19	1.22(0.90-1.67)
13	rs7991965	84804942	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.19	1.22(0.90-1.67)

13	rs114475966	84810618	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.19	1.22(0.90-1.67)
13	rs79673186	84812074	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.02	0.01	0.28	1.17(0.87-1.58)
13	rs7337340	84823667	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.02	0.01	0.33	1.15(0.85-1.55)
13	rs7338237	84824381	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.31	1.16(0.86-1.56)
13	rs115613124	84828092	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.21	1.21(0.89-1.65)
13	rs9742382	84830263	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.21	1.21(0.89-1.65)
13	rs75652849	84832487	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.4	1.14(0.83-1.56)
13	13:84838592:GAA:G	84838592	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	NA	NA	NA	NA
13	13:84840907:GA:G	84840907	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	NA	NA	NA	NA
13	rs79761693	84844398	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.17	1.23(0.90-1.67)
13	rs75702135	84849420	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.25	1.19(0.87-1.63)
13	rs77297708	84856250	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.21	1.21(0.89-1.65)
13	rs7322358	84861607	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.2	1.21(0.89-1.65)
13	rs7991421	84876709	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.2	1.21(0.89-1.65)
13	rs3904291	84886545	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.29	1.17(0.86-1.59)
13	rs7336054	84886948	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.29	1.17(0.86-1.59)
13	rs7336055	84886951	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.27	1.18(0.87-1.61)
13	rs80058242	84894022	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.27	1.19(0.87-1.62)
13	rs74102843	84897243	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.4	1.14(0.83-1.56)
13	rs59881804	84900226	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.4	1.14(0.83-1.56)
13	rs74102845	84900306	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.4	1.14(0.83-1.56)
13	rs60659114	84917204	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	NA	NA	NA	NA
13	rs1541073	84926232	0.032	0.008	3.6X10-06	5.08(2.48-10.42)	0.01	0.01	0.32	1.17(0.85-1.60)
5	5:669433_TG_T	669433	0.025	0.005	6.0X10-06	4.87(2.28-10.42)	0.01	0.009	0.01	1.51(1.08-2.10)
5	rs56934245	669442	0.025	0.005	6.0X10-06	4.87(2.28-10.42)	0.02	0.01	0.03	1.37(1.02-1.85)
10	rs117354281	60107698	0.047	0.018	8.4X10-06	2.76(1.73-4.39)	0.02	0.02	0.6	1.06(0.84-1.34)
5	rs62330305	674494	0.030	0.008	8.6X10-06	3.78(2.01-7.09)	0.02	0.01	0.01	1.46(1.08-1.96)
5	rs73044896	615195	0.066	0.030	1.1X10-05	2.28(1.56-3.32)	0.04	0.04	0.3	1.10(0.91-1.33)

CHR	Chromosome
SNP	Single nucleoitide polymorphism
BP	Base pair position
FCANOR	Allele freqency in Norwegian ME/CFS Cases
FCONOR	Allele freqency in Norwegian Controles
PNOR	P value Norwegian
OR(95%CI)NOR	Odds ratio with 95% confidence interval in Norwegian
FCAUK	Allele freqency in UK CFS Cases
FCOUK	Allele freqency in UK Controles
PUK	P value UK
OR(95%CI)UK	Odds ratio with 95% confidence interval in UK

Supplementary table 2. Association analyses on different subgroups of patients in the Norwegian discovery cohort

CHR	SNP	BP	AFall cases	AFcontrols	Pall cases	OR(CI95%)all cases	PInfection	AFInfection	Afnot infection	ORinfection	Pvaccination	AFvaccination	Afnot vaccination	ORvaccination	PAIDs	AFAIDs	AfnotAIDs	ORAIDs
2	rs7605661	43625078	0.3033	0.3714	0.0005061	0.7367(0.6-0.90)	0.001792	0.3013	0.3077	0.73	0.01915	0.2692	0.3094	0.6236	0.1305	0.3017	0.3035	0.7313
2	rs13011075	68639313	0.3267	0.2654	0.0009363	1.343(1.12-1.60)	0.0007303	0.3367	0.3038	1.405	0.2881	0.2231	0.3453	0.7946	0.1425	0.3276	0.3266	1.348
2	rs58089918	68639660	0.2963	0.2243	4.715x10-5	1.456(1.21-1.75)	8.926e-05	0.303	0.2808	1.504	0.2079	0.1769	0.3177	0.7435	0.08617	0.2931	0.2967	1.434
2	rs59998610	68646771	0.2986	0.2248	3.092e-05	1.468(1.22-1.76)	0.0001397	0.3013	0.2923	1.487	0.2034	0.1769	0.3204	0.7413	0.03332	0.3103	0.2967	1.552
2	rs7559803	181851177	0.2295	0.2963	0.0002729	0.7074(0.60-0.85)	0.0007796	0.2256	0.2385	0.6918	0.07538	0.2231	0.2307	0.6819	0.7865	0.2845	0.2209	0.9443
3	rs2582085	183488882	0.4227	0.3496	0.0002311	1.362(1.15-1.60)	1.488x10-5	0.4478	0.3654	1.508	0.9224	0.3538	0.4351	1.019	0.3042	0.3966	0.4268	1.222
5	rs3892626	543935	0.0363	0.01286	4.565x10-5	2.891(1.70-5.00)	0.0007696	0.03367	0.04231	2.675	0.002455	0.04615	0.03453	3.714	0.2401	0.02586	0.03794	2.038
5	rs16900931	620486	0.04684	0.02109	0.0001835	2.281(1.50-3.55)	0.0006346	0.04714	0.04615	2.296	0.06302	0.04615	0.04696	2.246	0.00648	0.06034	0.04472	2.981
5	rs115523291	672911	0.02459	0.004115	8.457x10-7	6.101(2.7-13.83)	1.957x10-6	0.02525	0.02308	6.269	0.003955	0.02308	0.02486	5.717	2.979x10-5	0.03448	0.02304	8.643
6	rs683571	165961530	0.05504	0.02726	0.0002684	2.078(1.40-3.10)	0.009305	0.04882	0.06923	1.831	0.02514	0.06154	0.05387	2.34	0.03956	0.06034	0.0542	2.291
7	rs2453836	45965795	0.2166	0.1553	8.259x10-5	1.504(1.22-1.85)	0.00122	0.2121	0.2269	1.464	0.07003	0.2154	0.2169	1.493	0.4599	0.181	0.2222	1.202
10	rs2590350	60014868	0.02108	0.006173	0.0004233	3.467(1.66-7.2)	0.07843	0.01347	0.03846	2.198	0.001914	0.03077	0.01934	5.111	0.0007408	0.03448	0.01897	5.75
13	rs7989859	84801379	0.03162	0.00823	3.643x10-6	3.934(2.1-7.3)	1.387x10-6	0.03535	0.02308	4.416	0.0109	0.03077	0.03177	3.825	0.05364	0.02586	0.03252	3.199
14	rs8017311	93105338	0.4532	0.3791	0.0002234	1.357(1.15-1.60)	0.003453	0.4461	0.4692	1.319	0.003571	0.5077	0.4434	1.689	0.005391	0.5086	0.4444	1.695
15	rs8028147	67418233	0.1452	0.2058	0.000156	0.6557(0.52-0.81)	0.02247	0.1633	0.1038	0.7534	0.8749	0.2	0.1354	0.965	0.3865	0.1724	0.1409	0.8042
15	rs12102184	67425143	0.1194	0.1692	0.0007781	0.6658(0.85-0.84)	0.04497	0.1347	0.08462	0.764	0.8212	0.1769	0.1091	1.055	0.3804	0.1379	0.1165	0.7854
19	rs281407	49233335	0.2646	0.338	0.0001214	0.705(0.60-0.85)	0.0005733	0.2626	0.2692	0.6977	0.5979	0.3154	0.2555	0.9024	0.05084	0.25	0.2669	0.653
19	rs8108136	49247693	0.4157	0.4907	0.0002485	0.7383(0.62-0.86)	0.0008238	0.4125	0.4231	0.7285	0.9724	0.4923	0.4019	1.006	0.4787	0.4569	0.4092	0.873
20	rs6089982	62401143	0.1967	0.1394	0.0001236	1.512(1.23-1.86)	0.00248	0.1902	0.2115	jan.45	0.01695	0.2154	0.1934	1.695	0.1323	0.1897	0.1978	1.445
22	rs9607658	39957740	0.5269	0.4594	0.0009861	1.311(1.11-1.54)	0.001924	0.532	0.5154	1.338	0.576	0.4846	0.5345	1.107	0.2245	0.5172	0.5285	1.261

Supplementary table 3. Stratified association analysis on clinical sub-groups in the Norwegian discovery cohort

CHR	SNP	AF cases	AF controls	P all cases	OR all cases	AF cases(infection)	AF cases(no-infection)	P infection vs no-infection	OR infection vs no-infection	AF cases AIDs	AF cases no-AIDs	P AIDs vs no-AIDs	OR AIDs vs no-AIDs	AF cases vaccine	AF cases no-vaccine	P vaccine vs no-vaccine	OR vaccine vs no-vaccine
2	rs7559803	0.2295	0.2963	0.00027 29	0.7074	0.2256	0.2252	0.9912	1.002	0.2845	0.2184	0.1161	1.423	0.2353	0.2169	0.6794	1.111
2	rs7605661	0.3033	0.3714	0.00050 61	0.7367	0.3013	0.3243	0.5268	0.8986	0.3017	0.3103	0.8524	0.9602	0.2941	0.3136	0.6952	0.9122
2	rs1301107 5	0.3267	0.2654	0.00093 63	1.343	0.3367	0.3018	0.3444	1.174	0.3276	0.329	0.9757	0.9935	0.2255	0.3492	0.01425	0.5427
2	rs5999861 0	0.2986	0.2248	3.092x1 0-5	1.468	0.3013	0.2928	0.8123	1.042	0.3103	0.2989	0.8026	1.056	0.1961	0.3186	0.01267	0.5215
2	rs5808991 8	0.2963	0.2243	4.715x1 0-5	1.456	0.303	0.2793	0.5086	1.122	0.2931	0.2989	0.9003	0.9728	0.1961	0.3186	0.01267	0.5215
3	rs2582085	0.4227	0.3496	0.00023 11	1.362	0.4478	0.3514	0.01302	1.497	0.3966	0.4296	0.505	0.8725	0.402	0.4339	0.5474	0.8769
5	rs1690093 1	0.04684	0.02109	0.00018 35	2.281	0.04714	0.04955	0.8857	0.9489	0.06034	0.04454	0.4556	1.378	0.04902	0.04407	0.8234	1.118
5	rs3892626	0.0363	0.01286	4.565x1 0-5	2.891	0.03367	0.04505	0.4422	0.7387	0.02586	0.03879	0.4942	0.6578	0.05882	0.03559	0.2632	1.693
5	rs1155232 91	0.02459	0.00411 5	8.457x1 0-7	6.101	0.02525	0.02252	0.8224	1.124	0.03448	0.02299	0.4596	1.518	0.01961	0.02373	0.7982	0.8229
6	rs683571	0.05504	0.02726	0.00026 84	2.078	0.04882	0.07658	0.1261	0.6189	0.06034	0.05316	0.7517	1.144	0.06863	0.05085	0.4611	1.375
7	rs2453836	0.2166	0.1553	8.259x1 0-5	1.504	0.2121	0.2252	0.6855	0.9262	0.181	0.2227	0.3133	0.7715	0.2353	0.222	0.7668	1.078
10	rs2590350	0.02108	0.00617 3	0.00042 33	3.467	0.01347	0.04505	0.00627 5	0.2894	0.03448	0.02011	0.3305	1.74	0.03922	0.01695	0.1402	2.367
13	rs7989859	0.03162	0.00823	3.643x1 0-6	3.934	0.03535	0.02252	0.353	1.591	0.02586	0.03305	0.6841	0.7768	0.02941	0.0339	0.8154	0.8636

14	rs8017311	0.4532	0.3791	0.000234	1.357	0.4461	0.464	0.6486	0.9306	0.5086	0.4397	0.1668	1.319	0.4314	0.4407	0.8612	0.9629
15	rs8028147	0.1452	0.2058	0.000156	0.6557	0.1633	0.1036	0.03214	1.689	0.1724	0.1422	0.3949	1.256	0.1569	0.1407	0.6664	1.136
15	rs12102184	0.1194	0.1692	0.000778	0.6658	0.1347	0.08108	0.03608	1.764	0.1379	0.1164	0.5076	1.215	0.1373	0.1136	0.4919	1.242
19	rs281407	0.2646	0.338	0.000121	0.705	0.2626	0.2838	0.5439	0.8989	0.25	0.2701	0.6504	0.9007	0.3235	0.261	0.1893	1.354
19	rs8108136	0.4157	0.4907	0.000248	0.7383	0.4125	0.4369	0.5283	0.9046	0.4569	0.4124	0.368	1.199	0.5	0.3966	0.05011	1.521
20	rs6089982	0.1967	0.1394	0.000123	1.512	0.1902	0.2207	0.3313	0.8294	0.1897	0.2011	0.7743	0.9295	0.2157	0.1983	0.6858	1.112
22	rs9607658	0.5269	0.4594	0.000986	1.311	0.468	0.482	0.7221	0.9455	0.4828	0.4713	0.8184	1.047	0.5	0.4661	0.5265	1.145

CHR	Chromosome
SNP	SNP
BP	Base pair position
AFcases	Allele frequency in the 427 Norwegian ME/CFS patients
AFcontrols	Allele frequency in the 972 controls
Pall cases	P value for the whole Norwegian cohort
ORall cases	Odds ratio for the whole Norwegian cohort
AFcases(infection)	Allele frequency in the Norwegian subgroup which reportedly developed ME/CFS after infectious disease (n=310)
AFcases(no-infection)	Allele frequency in the Norwegian subgroup which reportedly did NOT developed ME/CFS after infectious disease (n=117)
Pinfection vs no-infection	P value for the association between patients who reported disease development after Infectious onset and who did NOT
ORinfection vs no-infection	Odds ratio for the association between patients who reported disease development after Infectious onset and who did NOT
AFcases AIDs	Allele frequency in the Norwegian subgroup which have a comorbid autoimmune disorder (n=58)
AFcases no-AIDs	Allele frequency in the Norwegian subgroup which have NOT a comorbid autoimmune disorder (n=369)
PAIDs vs no-AIDs	P value for the association between patients who have a comorbid autoimmune disorder and who do NOT
ORAIDs vs no-AIDs	Odds ratio value for the association between patients who have a comorbid autoimmune disorder and who do NOT

AFcases vaccine	Allele frequency in the Norwegian subgroup which reportedly did developed ME/CFS after vaccination (n=65)
AFcases no-vaccine	Allele frequency in the Norwegian subgroup which reportedly did NOT developed ME/CFS after vaccination (n=362)
Pvaccine vs no-vaccine	P value for the association between patients who reported disease development after vaccination and who did NOT
ORvaccine vs no-vaccine	Odds ratio value for the association between patients who reported disease development after vaccination and who did NOT

Supplementary table 4. Replication of the tag SNPs in the Danish ME/CFS cohort and UK Biobank cohort

CHR	SNP	BP	FCA_{DK}	FCO_{DK}	P_{DK}	$OR(95\%CI)_{DK}$	FCA_{UK}	FCO_{UK}	P_{UK}	$OR(95CI)_{UK}$
2	rs7605661	43625078	0.3504	0.3623	0.5016	0.94(0.81-1.10)	0.3711	0.3804	0.3033	0.96(0.89-1.03)
2	rs13011075	68639313	0.3272	0.3018	0.1327	1.12(0.96-1.31)	0.3167	0.3076	0.2897	1.04(0.96-1.28)
2	rs58089918	68639660	0.2665	0.2611	0.7364	1.02(0.87-1.21)	0.2737	0.2706	0.7049	1.01(0.93-1.10)
2	rs59998610	68646771	0.2728	0.2664	0.6924	1.03(0.87-1.21)	0.2707	0.2674	0.6951	1.01(0.93-1.10)
2	rs7559803	181851177	0.2557	0.2659	0.5359	0.94(0.80-1.21)	0.2783	0.2736	0.575	1.02(0.94-1.11)
3	rs2582085	183488882	0.3498	0.3827	0.06455	0.86(0.74-1.00)	0.3629	0.3826	0.02784	0.91(0.85-0.99)
5	rs3892626	543935	0.01969	0.02112	0.7858	0.93(0.55-1.55)	0.0214	0.02146	0.9824	0.99(0.77-1.28)
5	rs16900931	620486	0.02609	0.02672	0.9148	0.97(0.62-1.52)	0.03707	0.0345	0.4519	1.07(0.88-1.38)
5	rs115523291	672911	0.01201	0.01374	0.6816	0.87(0.45-1.67)	0.01384	0.009535	0.02503	1.45(1.04-2.03)
6	rs683571	165961530	0.05111	0.03919	0.1053	1.31(0.94-1.85)	0.0378	0.03952	0.6322	0.95(0.79-1.15)
7	rs2453836	45965795	0.1736	0.2038	0.03951	0.82(0.67-0.99)	0.1895	0.1828	0.3513	1.04(0.95-1.14)
10	rs2590350	60014868	0.01751	0.01604	0.7526	1.09(0.62-1.90)	0.02352	0.02016	0.2081	1.17(0.91-1.49)
13	rs7989859	84801379	0.01747	0.01069	0.08925	1.64(0.92-2.94)	0.01496	0.01223	0.1926	1.22(0.90-1.67)
14	rs8017311	93105338	0.4052	0.3995	0.7494	1.02(0.88-1.86)	0.4278	0.4188	0.328	1.03(0.96-1.11)
15	rs8028147	67418233	0.2239	0.2112	0.3969	1.07(0.90-1.28)	0.2069	0.2054	0.8385	1.00(0.92-1.10)
15	rs12102184	67425143	0.1754	0.1623	0.3376	1.09(0.90-1.32)	0.1656	0.1698	0.536	0.96(0.87-1.06)
19	rs281407	49233335	0.3065	0.3163	0.5659	0.95(0.81-1.11)	0.3077	0.3177	0.2471	0.95(0.88-1.03)
19	rs8108136	49247693	0.4367	0.4392	0.8906	0.98(0.85-1.14)	0.4113	0.4297	0.04486	0.92(0.86-0.99)
20	rs6089982	62401143	0.1745	0.1585	0.2412	1.12(0.92-1.36)	0.1601	0.1466	0.04098	1.11(1.10-1.22)
22	rs9607658	39957740	0.4802	0.4977	0.3409	0.93(0.80-1.07)	0.4976	0.4988	0.8942	0.99(0.92-1.07)

CHR	Chromosome
SNP	Single nucleoitide polymorphism
BP	Base pair position
FCA_{DK}	Allele frequency Danish cases

FCO_{DK}	Allele frequency controls Danish controls
P_{DK}	P value in Danish data
$OR(95\%CI)_{DK}$	Odds ratio with 95% confidence interval in Danish data
FCA_{UK}	Allele frequency cases in the UK Biobank
FCO_{UK}	Allele frequency controls in the UK Biobank
P_{UK}	P value in the in the UK Biobank
OR(95%CI)UK	Odds ratio with 95% confidence interval in the UK Biobank

Supplementary table 5. Meta-analysis of the Norwegian, UK and Danish cohorts of the 20 putative ME/CFS tag SNPs revealed in the discovery screen

CHR	BP	SNP	Closest gene	Position	OR	P
20	62401143	rs6089982	*ZBTB46*	Intronic	1.16	0.0003
13	84801379	rs7989859	*LINC00333*	Intronic	1.56	0.0005
5	672911	rs115523291	*TPPP*	Intronic	1.57	0.002
19	49247693	rs8108136	*IZUMO1*	Intronic	0.91	0.002
2	68639313	rs13011075	*PLEK*	Intergenic	1.09	0.006
14	93105338	rs8017311	*RIN3*	Intronic	1.16	0.007
19	49233335	rs281407	*RASIP1*	Intronic	0.92	0.009
2	43625078	rs7605661	*THADA*	Intronic	0.92	0.02
10	60014868	rs2590350	*IPMK*	Intronic	1.27	0.03
2	68646771	rs59998610	*PLEK*	Intergenic	1.07	0.04
2	68639660	rs58089918	*PLEK*	Intergenic	1.07	0.05
5	620486	rs16900931	*CEP72*	Intronic	1.18	0.05
22	39957740	rs9607658	*CACNA1I*	Intergenic	0.95	0.08
6	165961530	rs683571	*PDE10A*	Intronic	1.14	0.09
3	183488882	rs2582085	*YEATS2*	Intronic	0.96	0.2
5	543935	rs3892626	*SLC9A3*	Intergenic	1.16	0.2
7	45965795	rs2453836	*IGFBP3*	Intergenic	1.06	0.2
15	67425143	rs12102184	*SMAD3*	Intronic	0.95	0.2
2	181851177	rs7559803	*UBE2E3*	Intronic	0.96	0.3
15	67418233	rs8028147	*SMAD3*	Intronic	0.97	0.5

CHR	Chromosome
BP	Base pair position
SNP	SNP RsID
Closest gene	Closest annotated gene
Position	Genomic location of the SNP
OR	Odds ratio
P	P value

Supplementary table 6. Association analyses in the UK Biobank dataset after filtering out illnesses that could potentially confound the diagnosis

CHR	SNP	BP	AFall cases	AFControls	AFcases no AIDs	AFcases no depression	AFcases primary	Pall cases	OR (CI95%)all cases	Pno AIDs	ORnoAIDs	Pno depresion	ORno depresion	Pprimary	ORprimary
2	rs7605661	43625078	0.3711	0.3804	0.3729	0.3723	0.3663	0.3033	0.96 (0.89-1.03)	0.4069	0.97	0.3969	0.9659	0.2781	0.9416
2	rs13011075	68639313	0.3167	0.3077	0.3152	0.3218	0.3118	0.2897	1.04 (0.96-1.13)	0.3896	1.036	0.1295	1.067	0.7343	01.feb
2	rs58089918	68639660	0.2737	0.2706	0.2718	0.2772	0.2743	0.7049	1.02 (0.93-1.10)	0.8861	1.006	0.4706	1.033	0.7542	1.019
2	rs59998610	68646771	0.2707	0.2675	0.269	0.2743	0.2709	0.6951	1.02 (0.93-1.10)	0.8539	1.008	0.4534	1.034	0.772	1.018
2	rs7559803	181851177	0.2783	0.2736	0.2765	0.2826	0.2709	0.575	1.02 (0.94-1.11)	0.7271	1.015	0.3249	1.045	0.8222	0.9864
3	rs2582085	183488882	0.3629	0.3826	0.3625	0.3654	0.3689	0.02784	0.92 (0.95-0.99)	0.02774	0.9179	0.07309	0.9292	0.2917	0.9433
5	rs3892626	543935	0.0214	0.02146	0.0218	0.02344	0.02427	0.9824	1.0 (0.77-1.28)	0.9001	1.016	0.4975	1.094	0.4706	1.134
5	rs16900931	620486	0.03707	0.03451	0.03764	0.03914	0.04005	0.4519	1.08 (0.88-1.3)	0.3669	1.094	0.209	1.139	0.2601	1.167
5	rs115523291	672911	0.01384	0.009537	0.01417	0.01464	0.01583	0.02503	1.46 (1.04-2.03)	0.01763	1.493	0.0132	1.542	0.02047	1.671
6	rs683571	165961530	0.0378	0.03953	0.03816	0.03745	0.03524	0.6322	0.95 (0.79-1.15)	0.7068	0.964	0.583	0.9448	0.4067	0.8877
7	rs2453836	45965795	0.1895	0.1828	0.1888	0.1874	0.1982	0.3513	1.05 (0.95-1.15)	0.413	1.04	0.5351	1.032	0.1385	1.105
10	rs2590350	60014868	0.02352	0.02017	0.02252	0.02513	0.02063	0.2081	1.17 (0.91-1.15)	0.3794	1.12	0.08414	1.252	0.9008	1.024
13	rs7989859	84801379	0.01496	0.01223	0.0146	0.01513	0.01517	0.1926	1.23 (0.90-1.70)	0.263	1.197	0.1958	1.24	0.3239	1.244
14	rs8017311	93105338	0.4278	0.4188	0.428	0.416	0.421	0.3284	1.037(0.96-1.11)	0.3228	1.038	0.8563	1.11	0.1338	1.063
15	rs8028147	67418233	0.2069	0.2055	0.2058	0.2117	0.2053	0.8385	1.01 (0.92-1.10)	0.9698	1.002	0.4403	1.038	0.9945	0.9995
15	rs12102184	67425143	0.1656	0.1699	0.1651	0.1683	0.1648	0.536	0.97 (0.90-1.1)	0.496	0.9662	0.8191	0.988	0.6172	0.9647
19	rs281407	49233335	0.3077	0.3175	0.3093	0.3126	0.315	0.2471	0.95 (0.90-1.03)	0.3451	0.9623	0.5762	0.9764	0.8317	0.9878
19	rs8108136	49247693	0.4113	0.4296	0.4148	0.4134	0.4262	0.04486	0.93 (0.86-1.00)	0.1115	0.9412	0.09709	0.9356	0.7919	0.9858

20	rs6089982	62401143	0.1601	0.1466	0.1616	0.1608	0.1748	0.04098	1.11 (1.00-1.30)	0.02529	1.122	0.04004	1.118	0.003139	1.233
22	rs9607658	39957740	0.4976	0.4988	0.5	0.496	0.5061	0.8942	1.00 (0.92-1.10)	0.9021	1.005	0.7807	0.989	0.5864	1.03

CHR	Chromosome
SNP	SNP rsID
BP	Base pair position
AFall cases	Allele frequency in the whole UK Biobank CFS cohort (N=2105)
AFControls	Allele frequency in the control data (N=4876)
AFcases no AIDs	Allele frequency of the CFS cohort after excluding patients with AIDs (excluded N=85)
AFcases no depression	Allele frequency of the CFS cohort after excluding patients with depression (excluded N=354)
AFcases primary	Allele frequency of the CFS cohort after only including patients who have CFS as first entry in the Biobank (excluded N=1281)
Pall cases	P value of the whole UK Biobank CFS cohort (N=2105)
OR (CI95%)all cases	Odds ratio (confidence interval)of the whole UK Biobank CFS cohort (N= 2105)
Pno AIDs	P value of the whole UK Biobank CFS cohort with exluded autoimmune disorders (excluded N=85)
ORnoAIDs	Odds ratio of the whole UK Biobank CFS cohort with exluded autoimmune disorders (excluded N=85)
Pno depresion	P value of the whole UK Biobank CFS cohort with exluded depression (excluded N=354)
ORno depresion	Odds ratio of the whole UK Biobank CFS cohort with exluded depression (excluded N=354)
Pprimary	P value of the UK Biobank CFS cohort with only included patietns who have CFS as a first entry (excluded N= 1281)
ORprimary	Odds ratio of the UK Biobank CFS cohort with only included patietns who have CFS as a first entry (excluded N=1281)

Supplementary Table 7. SNPs from the UK Biobank with P<1x10-5 and data from the Norwegian cohort for the overlapping SNPs

CHR	SNP	BP	FCAUK	FCOUK	PUK	ORUK	FCANOR	FCONOR	PNOR	OR(95%CI)NOR
6	**rs193219461**	**90947107**	**0.02057**	**0.01082**	**6.662x10-06**	**1.92(1.43-2.56)**	**0.02342**	**0.01749**	**0.2938**	**1.34(0.77-2.35)**
13	**rs9526104**	**46295836**	**0.08043**	**0.1057**	**5.224x10-06**	**0.7399(0.64-0.84)**	**0.04098**	**0.03652**	**0.5692**	**1.12(0.74-1.70)**
6	**rs72914217**	**94006318**	**0.137**	**0.1093**	**4.782x10-06**	**1.294(1.15-1.44)**	**0.09016**	**0.07305**	**0.1205**	**1.25(0.94-1.68)**
6	**rs17448051**	**94005099**	**0.1393**	**0.1121**	**8.747x10-06**	**1.282(1.14-1.43)**	**0.09016**	**0.07305**	**0.1205**	**1.25(0.94-1.68)**
13	**rs11147812**	**41404706**	**0.4494**	**0.4912**	**7.707x10-06**	**0.8455(0.78-0.91)**	**0.3934**	**0.4115**	**0.3698**	**0.92(0.78-1.09)**
22	**rs1858756**	**51105101**	**0.5067**	**0.4646**	**5.472x10-06**	**1.183(1.10-1.27)**	**0.4731**	**0.4882**	**0.4617**	**0.94(0.80-1.10)**
22	**rs9616812**	**51105556**	**0.5064**	**0.4652**	**8.382x10-06**	**1.18(1.09-1.26)**	**0.4731**	**0.4882**	**0.4617**	**0.94(0.80-1.10)**
22	**rs9616906**	**51104680**	**0.4708**	**0.43**	**9.008x10-06**	**1.18(1.09-1.26)**	**0.4789**	**0.4918**	**0.5313**	**0.94(0.80-1.11)**
22	**rs752183198**	**51106681**	**0.4708**	**0.4299**	**9.111x10-06**	**1.18(1.09-1.26)**	**0.4813**	**0.4964**	**0.4609**	**0.94(0.80-1.10)**
22	**rs11704245**	**51089624**	**0.4571**	**0.4146**	**6.336x10-06**	**1.189(1.10-1.28)**	**0.4836**	**0.4676**	**0.4346**	**1.06(0.90-1.25)**
22	**rs9616900**	**51098443**	**0.4978**	**0.456**	**7.396x10-06**	**1.183(1.09-1.27)**	**0.4848**	**0.5021**	**0.3999**	**0.93(0.79-1.09)**
18	rs79110322	40386803	0.01948	0.008671	7.756x10-08	2.27(1.67-3.08)	NA	NA	NA	NA
3	rs76284084	1,96E+08	0.02314	0.01237	2.887x10-06	1.891(1.44-2.47)	NA	NA	NA	NA
17	rs2017476	62067638	0.02748	0.01503	9.02x10-07	1.852(1.44-2.37)	NA	NA	NA	NA
22	rs116943399	44805222	0.02555	0.01433	5.078x10-06	1.804(1.39-2.33)	NA	NA	NA	NA
8	rs150846691	66582741	0.02618	0.01493	6.078x10-06	1.774(1.37-2.28)	NA	NA	NA	NA
4	rs146794331	43914433	0.02548	0.01458	9.892x10-06	1.767(1.36-2.28)	NA	NA	NA	NA
8	rs141516949	66565126	0.02618	0.01513	9.702x10-06	1.75(1.36-2.24)	NA	NA	NA	NA
3	rs2432533	1,96E+08	0.03591	0.02223	6.845x10-06	1.638(1.31-2.03	NA	NA	NA	NA
3	rs75692609	1,96E+08	0.03982	0.02518	5.41x10-06	1.606(1.30-1.97)	NA	NA	NA	NA
3	rs11923648	1,96E+08	0.04109	0.02643	4.635x10-06	1.578(1.29-1.92)	NA	NA	NA	NA
2	rs114846046	1,78E+08	0.04765	0.03114	2.127x10-06	1.557(1.29-1.87)	NA	NA	NA	NA
17	rs9898354	71827146	0.055	0.03746	3.245x10-06	1.495(1.26-1.77)	NA	NA	NA	NA
17	rs150214709	71831860	0.07846	0.05674	2.178x10-06	1.415(1.22-1.63)	NA	NA	NA	NA

17	rs150606845	71829764	0.08182	0.06075	7.758x10-06	1.378(1.19-1.58)	NA	NA	NA	NA
11	rs575274850	62649960	0.1113	0.0847	2.348x10-06	1.353(1.19-1.53)	NA	NA	NA	NA
11	rs489381	62652063	0.1111	0.0846	9.392x10-07	1.352(1.19-1.52)	NA	NA	NA	NA
8	rs112954458	8275605	0.1018	0.07781	5.555x10-06	1.343(1.18-1.52)	NA	NA	NA	NA
20	rs73101112	24299978	0.1052	0.08167	9.535x10-06	1.322(1.16-1.49)	NA	NA	NA	NA
6	rs113946226	93999192	0.1687	0.1367	1.12x10-06	1.281(1.15-1.41)	NA	NA	NA	NA
4	rs1531713	1,59E+08	0.1889	0.1545	1.162x10-06	1.274(1.15-1.40)	NA	NA	NA	NA
6	rs17447037	93962762	0.1691	0.1379	2.038x10-06	1.273(1.15-1.40)	NA	NA	NA	NA
6	rs3799821	93962293	0.1691	0.1379	2.087x10-06	1.272(1.15-1.40)	NA	NA	NA	NA
6	rs17446388	93947347	0.1698	0.1386	2.125x10-06	1.272(1.15-1.40)	NA	NA	NA	NA
6	rs77585483	93926046	0.1695	0.1384	2.156x10-06	1.271(1.15-1.40)	NA	NA	NA	NA
6	rs73755355	93959771	0.1692	0.1381	2.279x10-06	1.271(1.15-1.40)	NA	NA	NA	NA
6	rs58136974	93983205	0.1688	0.1378	2.35x10-06	1.271(1.15-1.40)	NA	NA	NA	NA
6	rs60964062	93981188	0.1688	0.1378	2.35x10-06	1.271(1.15-1.40)	NA	NA	NA	NA
6	rs73530353	93980350	0.1688	0.1378	2.35x10-06	1.271(1.15-1.40)	NA	NA	NA	NA
6	rs73755360	93983872	0.1688	0.1378	2.35x10-06	1.271(1.15-1.40)	NA	NA	NA	NA
6	rs78307356	93979780	0.1688	0.1378	2.35x10-06	1.271(1.15-1.40)	NA	NA	NA	NA
6	rs57565122	93971567	0.169	0.138	2.408x10-06	1.271(1.15-1.40)	NA	NA	NA	NA
6	rs61145076	93971964	0.169	0.138	2.408x10-06	1.271(1.15-1.40)	NA	NA	NA	NA
6	rs17446661	93957161	0.1693	0.1384	2.711x10-06	1.269(1.14-1.40)	NA	NA	NA	NA
6	rs17543546	93947670	0.1692	0.1384	2.797x10-06	1.268(1.14-1.40)	NA	NA	NA	NA
5	rs142519313	18542863	0.158	0.1292	6.44x10-06	1.265(1.14-1.40)	NA	NA	NA	NA
5	rs79188774	18546131	0.1582	0.1296	7.42x10-06	1.263(1.14-1.39)	NA	NA	NA	NA
5	rs78909576	18563476	0.1564	0.1281	9.44x10-06	1.262(1.13-1.39)	NA	NA	NA	NA
5	rs76847610	18552120	0.1576	0.1292	8.8x10-06	1.261(1.13-1.39)	NA	NA	NA	NA
5	rs76505864	18566421	0.1567	0.1284	9.94x10-06	1.261(1.13-1.39)	NA	NA	NA	NA
12	rs112473595	95855104	0.2565	0.2158	3.422x10-07	1.25(1.14-1.36)	NA	NA	NA	NA
19	19:52128217_CA_C	52128217	0.1747	0.1449	9.159x10-06	1.249(1.13-1.37)	NA	NA	NA	NA
17	rs2044165	47761682	0.2068	0.1738	4.394x10-06	1.24(1.31-1.35)	NA	NA	NA	NA

6	6:22368249_TA_T	22368249	0.2242	0.1896	4.87x10-06	1.235(1.12-1.35)	NA	NA	NA	NA
3	rs62296617	1,77E+08	0.2072	0.1753	9.955x10-06	1.23(1.12-1.34)	NA	NA	NA	NA
17	rs66517465	47967360	0.2323	0.1983	7.974x10-06	1.223(1.12-1.33)	NA	NA	NA	NA
5	rs7729710	18562868	0.3376	0.2953	8.8x10-07	1.216(1.25-1.31)	NA	NA	NA	NA
10	10:21318908_ACGTCCCC_A	21318908	0.413	0.367	3.584x10-07	1.21(1.26-1.30)	NA	NA	NA	NA
10	rs638929	21318625	0.4142	0.3687	4.498x10-07	1.21(1.24-1.30)	NA	NA	NA	NA
10	rs3858203	21318457	0.4521	0.4049	2.852x10-07	1.21(1.12-1.30)	NA	NA	NA	NA
10	rs640729	21319000	0.4142	0.3689	5.25x10-07	1.21(1.12-1.30)	NA	NA	NA	NA
12	rs770947392	1432336	0.2875	0.2505	8.017x10-06	1.207(1.11-1.31)	NA	NA	NA	NA
10	rs643674	21338595	0.4159	0.3718	1.002x10-06	1.203(1.11-1.29)	NA	NA	NA	NA
10	rs623047	21323397	0.4149	0.3712	1.189x10-06	1.202(1.16-1.29)	NA	NA	NA	NA
10	rs623551	21323240	0.4149	0.3712	1.189x10-06	1.202(1.16-1.29)	NA	NA	NA	NA
10	rs113117610	21322682	0.415	0.3712	1.199x10-06	1.202(1.16-1.29)	NA	NA	NA	NA
10	rs499364	21337917	0.4156	0.3719	1.214x10-06	1.202(1.16-1.29)	NA	NA	NA	NA
10	rs589425	21322443	0.4148	0.3711	1.218x10-06	1.202(1.16-1.29)	NA	NA	NA	NA
10	rs590763	21322756	0.4148	0.3711	1.218x10-06	1.202(1.16-1.29)	NA	NA	NA	NA
10	rs638548	21322167	0.4148	0.3711	1.218x10-06	1.202(1.16-1.29)	NA	NA	NA	NA
10	rs684390	21321890	0.4148	0.3711	1.218x10-06	1.202(1.16-1.29)	NA	NA	NA	NA
10	rs685192	21322030	0.4148	0.3711	1.218x10-06	1.202(1.16-1.29)	NA	NA	NA	NA
10	rs518712	21324908	0.4156	0.3719	1.259x10-06	1.201(1.11-1.29)	NA	NA	NA	NA
10	rs683469	21321685	0.4148	0.3712	1.274x10-06	1.201(1.11-1.29)	NA	NA	NA	NA
10	rs2119240	21344931	0.4156	0.372	1.281x10-06	1.201(1.11-1.29)	NA	NA	NA	NA
10	rs652864	21319399	0.4148	0.3696	5.298x10-07	1.20(1.12-1.30)	NA	NA	NA	NA
10	rs640750	21319010	0.4142	0.3689	5.316x10-07	1.20(1.12-1.30)	NA	NA	NA	NA
10	rs67179	21318892	0.4151	0.3699	5.923x10-07	1.20(1.12-1.30)	NA	NA	NA	NA
10	rs7918689	21317451	0.4561	0.4099	6.125x10-07	1.20(1.12-1.30)	NA	NA	NA	NA
10	rs7898342	21317504	0.4561	0.4101	7.025x10-07	1.20(1.12-1.29)	NA	NA	NA	NA
10	rs613407	21337794	0.4154	0.3719	1.344x10-06	1.2(1.11-1.29)	NA	NA	NA	NA
10	rs634614	21325698	0.4156	0.3721	1.345x10-06	1.2(1.11-1.29)	NA	NA	NA	NA

10	rs603346	21342448	0.4154	0.3719	1.383x10-06	1.2(1.11-1.29)	NA	NA	NA	NA
10	rs7079238	21346197	0.4152	0.3718	1.387x10-06	1.2(1.11-1.29)	NA	NA	NA	NA
10	rs604251	21323423	0.4146	0.3712	1.394x10-06	1.2(1.11-1.29)	NA	NA	NA	NA
10	rs620853	21324820	0.4156	0.3721	1.397x10-06	1.2(1.11-1.29)	NA	NA	NA	NA
10	rs633714	21332019	0.4154	0.3718	1.3x10-06	1.2(1.11-1.29)	NA	NA	NA	NA
10	rs594599	21342506	0.4154	0.372	1.403x10-06	1.2(1.11-1.29)	NA	NA	NA	NA
10	rs16921489	21343578	0.4154	0.372	1.423x10-06	1.2(1.11-1.29)	NA	NA	NA	NA
10	rs2165286	21345383	0.4154	0.372	1.423x10-06	1.2(1.11-1.29)	NA	NA	NA	NA
10	rs4025869	21345352	0.4154	0.372	1.423x10-06	1.2(1.11-1.29)	NA	NA	NA	NA
10	rs543011	21343413	0.4154	0.372	1.423x10-06	1.2(1.11-1.29)	NA	NA	NA	NA
10	rs694004	21343427	0.4154	0.372	1.423x10-06	1.2(1.11-1.29)	NA	NA	NA	NA
10	rs386741753	21326753	0.4164	0.3729	1.489x10-06	1.2(1.11-1.29)	NA	NA	NA	NA
10	rs7077775	21345971	0.4152	0.3719	1.514x10-06	1.199(1.11-1.29)	NA	NA	NA	NA
10	rs568929	21342831	0.4154	0.3721	1.534x10-06	1.199(1.11-1.29)	NA	NA	NA	NA
10	rs561415	21319705	0.4148	0.3715	1.544x10-06	1.199(1.11-1.29)	NA	NA	NA	NA
10	rs35775412	21324666	0.4157	0.3724	1.546x10-06	1.199(1.11-1.29)	NA	NA	NA	NA
10	rs649734	21326720	0.4156	0.3724	1.561x10-06	1.199(1.11-1.29)	NA	NA	NA	NA
10	rs494226	21325247	0.4163	0.373	1.564x10-06	1.199(1.11-1.29)	NA	NA	NA	NA
10	rs682570	21321492	0.4148	0.3716	1.592x10-06	1.199(1.11-1.29)	NA	NA	NA	NA
10	rs654717	21319832	0.4148	0.3716	1.615x10-06	1.199(1.11-1.29)	NA	NA	NA	NA
10	rs671173	21321296	0.4148	0.3716	1.615x10-06	1.199(1.11-1.29)	NA	NA	NA	NA
10	rs589483	21329049	0.4154	0.3722	1.623x10-06	1.199(1.11-1.29)	NA	NA	NA	NA
10	rs4748745	21344463	0.4152	0.372	1.6x10-06	1.199(1.11-1.29)	NA	NA	NA	NA
10	rs7078477	21346392	0.4156	0.3724	1.6x10-06	1.199(1.11-1.29)	NA	NA	NA	NA
10	rs2119241	21344875	0.4151	0.372	1.666x10-06	1.198(1.11-1.29)	NA	NA	NA	NA
10	rs534865	21326895	0.4154	0.3722	1.67x10-06	1.198(1.11-1.29)	NA	NA	NA	NA
10	rs534149	21326934	0.4154	0.3723	1.746x10-06	1.198(1.11-1.29)	NA	NA	NA	NA
10	rs641632	21329425	0.4141	0.3711	1.758x10-06	1.198(1.11-1.29)	NA	NA	NA	NA
10	rs4747430	21344294	0.4152	0.3722	1.803x10-06	1.198(1.11-1.29)	NA	NA	NA	NA

10	rs1543831	21345712	0.415	0.372	1.808x10-06	1.198(1.11-1.29)	NA	NA	NA	NA
10	rs672500	21327127	0.4154	0.3724	1.825x10-06	1.198(1.11-1.29)	NA	NA	NA	NA
10	10:21324255_TC_T	21324255	0.4131	0.3701	1.827x10-06	1.198(1.11-1.29)	NA	NA	NA	NA
10	rs1543829	21345889	0.4136	0.3707	1.873x10-06	1.198(1.11-1.29)	NA	NA	NA	NA
10	rs656069	21320162	0.4156	0.3727	2.083x10-06	1.197(1.11-1.28)	NA	NA	NA	NA
10	rs602053	21342703	0.4154	0.3727	2.144x10-06	1.196(1.11-1.28)	NA	NA	NA	NA
10	rs2119239	21345220	0.4154	0.3727	2.149x10-06	1.196(1.11-1.28)	NA	NA	NA	NA
10	rs7901012	21348358	0.4154	0.3727	2.274x10-06	1.196(1.11-1.28)	NA	NA	NA	NA
10	rs7911561	21348098	0.4154	0.3727	2.274x10-06	1.196(1.11-1.28)	NA	NA	NA	NA
10	10:21326371_GA_G	21326371	0.4144	0.3718	2.277x10-06	1.196(1.11-1.28)	NA	NA	NA	NA
10	rs508282	21327519	0.4155	0.3729	2.355x10-06	1.195(1.11-1.28)	NA	NA	NA	NA
10	rs665486	21327969	0.4159	0.3733	2.368x10-06	1.195(1.11-1.28)	NA	NA	NA	NA
10	rs11012498	21347472	0.4154	0.3728	2.371x10-06	1.195(1.11-1.28)	NA	NA	NA	NA
10	rs4748746	21349080	0.415	0.3726	2.499x10-06	1.195(1.10-1.28)	NA	NA	NA	NA
10	rs663759	21327593	0.4155	0.373	2.511x10-06	1.195(1.10-1.28)	NA	NA	NA	NA
10	rs10828179	21348863	0.415	0.3726	2.535x10-06	1.195(1.10-1.28)	NA	NA	NA	NA
10	rs11012499	21347532	0.4151	0.3726	2.5x10-06	1.195(1.10-1.28)	NA	NA	NA	NA
10	rs678185	21335103	0.417	0.3746	2.583x10-06	1.194(1.10-1.28)	NA	NA	NA	NA
10	rs2986673	21332250	0.417	0.3746	2.613x10-06	1.194(1.10-1.28)	NA	NA	NA	NA
10	rs893882	21349172	0.4151	0.3727	2.617x10-06	1.194(1.10-1.28)	NA	NA	NA	NA
10	rs773253512	21327425	0.4163	0.3739	2.641x10-06	1.194(1.10-1.28)	NA	NA	NA	NA
10	10:21345340_CA_C	21345340	0.4148	0.3725	2.763x10-06	1.194(1.10-1.28)	NA	NA	NA	NA
10	rs893881	21349368	0.4151	0.3729	2.896x10-06	1.193(1.10-1.28)	NA	NA	NA	NA
10	rs4747431	21350153	0.414	0.3719	3.066x10-06	1.193(1.10-1.28)	NA	NA	NA	NA
10	rs611992	21337438	0.417	0.3752	3.452x10-06	1.191(1.10-1.28)	NA	NA	NA	NA
10	10:21316230_TA_T	21316230	0.4554	0.4125	7.517x10-06	1.191(1.10-1.28)	NA	NA	NA	NA
10	rs491386	21341618	0.417	0.3755	4.204x10-06	1.19(1.10-1.28)	NA	NA	NA	NA
10	rs561753	21329087	0.417	0.3757	4.822x10-06	1.19(1.10-1.28)	NA	NA	NA	NA
10	rs757399260	21321562	0.4187	0.377	4.822x10-06	1.19(1.10-1.28)	NA	NA	NA	NA

10	rs2119243	21344747	0.4177	0.3764	4.514x10-06	1.189(1.10-1.28)	NA	NA	NA	NA
10	rs613306	21337716	0.4175	0.3762	4.675x10-06	1.188(1.10-1.28)	NA	NA	NA	NA
10	rs2986672	21327294	0.4169	0.3755	4.761x10-06	1.188(1.10-1.28)	NA	NA	NA	NA
10	rs703106	21327276	0.4165	0.3753	5.274x10-06	1.188(1.10-1.28)	NA	NA	NA	NA
10	rs603790	21342348	0.4177	0.3765	4.792x10-06	1.188(1.10-1.27)	NA	NA	NA	NA
10	rs587211	21328563	0.417	0.3758	5.101x10-06	1.188(1.10-1.27)	NA	NA	NA	NA
10	rs669320	21327858	0.4165	0.3753	4.892x10-06	1.188(1.03-1.27)	NA	NA	NA	NA
10	rs631361	21336002	0.4175	0.3764	5.078x10-06	1.188(1.03-1.27)	NA	NA	NA	NA
10	rs579249	21322858	0.417	0.3759	5.471x10-06	1.187(1.10-1.27)	NA	NA	NA	NA
10	rs4025868	21345007	0.4177	0.3767	5.503x10-06	1.187(1.10-1.27)	NA	NA	NA	NA
10	rs525350	21340179	0.4175	0.3765	5.593x10-06	1.187(1.10-1.27)	NA	NA	NA	NA
10	rs1543830	21345825	0.4185	0.3775	5.703x10-06	1.187(1.10-1.27)	NA	NA	NA	NA
10	rs10828183	21353551	0.414	0.3731	5.746x10-06	1.187(1.10-1.27)	NA	NA	NA	NA
10	rs674276	21333311	0.4175	0.3766	5.97x10-06	1.187(1.10-1.27)	NA	NA	NA	NA
10	rs2165287	21327295	0.4167	0.3757	5.987x10-06	1.187(1.10-1.27)	NA	NA	NA	NA
10	rs495866	21331682	0.4175	0.3767	6.149x10-06	1.187(1.10-1.27)	NA	NA	NA	NA
1	rs2673978	2,35E+08	0.4038	0.4446	9.947x10-06	0.846(0.78-0.91)	NA	NA	NA	NA
1	rs2459153	2,35E+08	0.4024	0.4445	5.506x10-06	0.8414(0.78-0.90)	NA	NA	NA	NA
11	rs7129155	16934130	0.3791	0.4211	5.25x10-06	0.8396(0.77-0.90)	NA	NA	NA	NA
1	rs6698918	2,35E+08	0.4014	0.4446	3.342x10-06	0.8378(0.77-0.90)	NA	NA	NA	NA
4	rs4591549	1,61E+08	0.2956	0.3341	9.56x10-06	0.8365(0.77-0.90)	NA	NA	NA	NA
4	rs7674279	1,61E+08	0.2923	0.3305	1x10-05	0.8364(0.77-0.90)	NA	NA	NA	NA
4	rs9996139	1,61E+08	0.294	0.3327	8.482x10-06	0.8354(0.77-0.90)	NA	NA	NA	NA
2	rs4578869	1,32E+08	0.2734	0.3109	9.646x10-06	0.834(0.76-0.90)	NA	NA	NA	NA
2	rs4850271	1,32E+08	0.2736	0.3111	9.681x10-06	0.834(0.76-0.90)	NA	NA	NA	NA
4	rs750154932	1,61E+08	0.2932	0.3322	7.678x10-06	0.8339(0.77-0.90)	NA	NA	NA	NA
2	rs4850130	1,32E+08	0.2735	0.3111	9.419x10-06	0.8338(0.76-0.90)	NA	NA	NA	NA
2	rs113244020	1,32E+08	0.2728	0.3105	9.428x10-06	0.8338(0.76-0.90)	NA	NA	NA	NA
12	rs11114545	81043969	0.1924	0.2271	9.217x10-06	0.811(0.73-0.88)	NA	NA	NA	NA

18	rs1877055	56656078	0.1497	0.1808	8.88x10-06	0.7978(0.72-0.88)	NA	NA	NA	NA
16	rs534562792	20741105	0.1322	0.1658	1.646x10-06	0.7666(0.68-0.85)	NA	NA	NA	NA
4	rs113877996	66138619	0.08764	0.1136	5.281x10-06	0.7493(0.66-0.84)	NA	NA	NA	NA
4	rs75457531	66057362	0.08221	0.107	8.532x10-06	0.7477(0.65-0.85)	NA	NA	NA	NA
4	rs74628869	65947122	0.08632	0.1124	4.411x10-06	0.7463(0.65-0.84)	NA	NA	NA	NA
4	4:65938593_GA_G	65938593	0.08557	0.1117	4.805x10-06	0.7443(0.65-0.84)	NA	NA	NA	NA
14	rs11847006	29977539	0.06202	0.08476	5.205x10-06	0.7139(0.61-0.82)	NA	NA	NA	NA
14	rs66490311	29981357	0.05854	0.08019	7.485x10-06	0.7132(0.61-0.82)	NA	NA	NA	NA
14	rs73251477	29981137	0.05854	0.08019	7.485x10-06	0.7132(0.61-0.82)	NA	NA	NA	NA
14	rs17114871	29979967	0.05833	0.08002	7.099x10-06	0.7122(0.61-0.82)	NA	NA	NA	NA
20	rs142964241	20874741	0.05639	0.07771	9.328x10-06	0.7093(0.60-0.82)	NA	NA	NA	NA
14	rs750483590	29981186	0.05858	0.08098	4.122x10-06	0.7062(0.60-0.81)	NA	NA	NA	NA
6	rs77148132	1,56E+08	0.04133	0.06101	3.04x10-06	0.6635(0.55-0.78)	NA	NA	NA	NA
6	rs57304622	1,56E+08	0.04135	0.06113	2.775x10-06	0.6625(0.55-0.78)	NA	NA	NA	NA
6	rs78305618	1,56E+08	0.04109	0.0608	2.818x10-06	0.6619(0.55-0.78)	NA	NA	NA	NA
6	rs78534173	1,56E+08	0.04109	0.06091	2.532x10-06	0.6607(0.55-0.78)	NA	NA	NA	NA
6	rs13362645	1,56E+08	0.04087	0.06064	2.555x10-06	0.6601(0.55-0.78)	NA	NA	NA	NA
6	rs759167457	1,56E+08	0.04095	0.06128	1.486x10-06	0.6541(0.54-0.77)	NA	NA	NA	NA
6	rs9480197	1,56E+08	0.03709	0.0564	1.873x10-06	0.6444(0.53-0.77)	NA	NA	NA	NA
6	rs144040479	1,56E+08	0.03661	0.05649	9.184x10-07	0.6348(0.52-0.76)	NA	NA	NA	NA
6	rs9397249	1,56E+08	0.02755	0.04418	3.535x10-06	0.613(0.49-0.75)	NA	NA	NA	NA
8	rs74399940	88610057	0.01954	0.03337	9.139x10-06	0.5774(0.45-0.73)	NA	NA	NA	NA
6	rs117395841	1,56E+08	0.02052	0.03619	1.231x10-06	0.5577(0.43-0.70)	NA	NA	NA	NA
6	rs78235135	1,56E+08	0.02052	0.03619	1.231x10-06	0.5577(0.43-0.70)	NA	NA	NA	NA
6	rs117163443	1,56E+08	0.02052	0.03628	1.094x10-06	0.5563(0.43-0.70)	NA	NA	NA	NA
12	rs74553645	1,15E+08	0.01617	0.029	9.216x10-06	0.5502(0.42-0.71)	NA	NA	NA	NA
6	rs118110730	1,56E+08	0.02029	0.03621	8.243x10-07	0.55(0.43-0.70)	NA	NA	NA	NA
12	rs143565916	1,15E+08	0.01618	0.0292	7.151x10-06	0.5466(0.41-0.71)	NA	NA	NA	NA
12	rs115307297	1,15E+08	0.01592	0.02878	7.984x10-06	0.5459(0.41-0.71)	NA	NA	NA	NA

12	rs112003527	1,15E+08	0.01592	0.02889	6.886x10-06	0.5438(0.41-0.71)	NA	NA	NA	NA
12	rs113762292	1,15E+08	0.01592	0.02889	6.886x10-06	0.5438(0.41-0.71)	NA	NA	NA	NA
12	rs116266993	1,15E+08	0.01592	0.02889	6.886x10-06	0.5438(0.41-0.71)	NA	NA	NA	NA
12	rs77755486	1,15E+08	0.01592	0.02889	6.886x10-06	0.5438(0.41-0.71)	NA	NA	NA	NA
12	rs114643729	1,15E+08	0.01592	0.02889	6.939x10-06	0.5438(0.41-0.71)	NA	NA	NA	NA
12	rs75968991	1,15E+08	0.01593	0.02899	6.181x10-06	0.5423(0.41-0.70)	NA	NA	NA	NA
12	rs80033619	1,15E+08	0.01593	0.02899	6.181x10-06	0.5423(0.41-0.70)	NA	NA	NA	NA
12	rs111352345	1,15E+08	0.01592	0.02898	6.122x10-06	0.5421(0.41-0.70)	NA	NA	NA	NA
12	rs113156850	1,15E+08	0.01592	0.02898	6.122x10-06	0.5421(0.41-0.70)	NA	NA	NA	NA
12	rs112943332	1,15E+08	0.01592	0.029	5.981x10-06	0.5418(0.41-0.70)	NA	NA	NA	NA
12	rs117428622	1,15E+08	0.01593	0.0291	5.326x10-06	0.5401(0.41-0.70)	NA	NA	NA	NA
6	rs74930209	1,56E+08	0.0205	0.0369	4.67x10-07	0.54(0.43-0.69)	NA	NA	NA	NA
12	rs112908253	1,15E+08	0.01624	0.02986	3.404x10-06	0.5362(0.41-0.70)	NA	NA	NA	NA
12	rs116255258	1,15E+08	0.01569	0.02888	4.668x10-06	0.536(0.40-0.70)	NA	NA	NA	NA
12	rs112174785	1,15E+08	0.01569	0.02899	4.048x10-06	0.534(0.40-0.70)	NA	NA	NA	NA
12	rs75344510	1,15E+08	0.01569	0.02899	4.048x10-06	0.534(0.40-0.70)	NA	NA	NA	NA
12	rs115745341	1,15E+08	0.01569	0.02898	4.079x10-06	0.534(0.40-0.70)	NA	NA	NA	NA
12	rs118132434	1,15E+08	0.01569	0.02898	4.079x10-06	0.534(0.40-0.70)	NA	NA	NA	NA
12	rs189603264	1,15E+08	0.01568	0.02916	3.074x10-06	0.5303(0.40-0.69)	NA	NA	NA	NA
12	rs112016583	1,15E+08	0.01569	0.02918	3.087x10-06	0.5303(0.40-0.69)	NA	NA	NA	NA
12	rs113399068	1,15E+08	0.01569	0.02918	3.087x10-06	0.5303(0.40-0.69)	NA	NA	NA	NA
12	rs111753839	1,15E+08	0.01568	0.02926	2.683x10-06	0.5284(0.40-0.69)	NA	NA	NA	NA
12	rs75200220	1,15E+08	0.01568	0.02926	2.683x10-06	0.5284(0.40-0.69)	NA	NA	NA	NA
12	rs75954846	1,15E+08	0.01568	0.02926	2.683x10-06	0.5284(0.40-0.69)	NA	NA	NA	NA
12	rs113928249	1,15E+08	0.01568	0.02936	2.34x10-06	0.5266(0.40-0.69)	NA	NA	NA	NA
12	rs112362286	1,15E+08	0.01473	0.02791	3.027x10-06	0.5206(0.39-0.68)	NA	NA	NA	NA
6	rs117579163	1,56E+08	0.01931	0.03644	1.057x10-07	0.52(0.40-0.66)	NA	NA	NA	NA
6	rs76696149	1,56E+08	0.0205	0.03752	2.031x10-07	0.52(0.40-0.66)	NA	NA	NA	NA
6	rs150884129	1,56E+08	0.01739	0.03414	7.155x10-08	0.50(0.38-0.64)	NA	NA	NA	NA

6	rs77381650	1,56E+08	0.01713	0.03411	4.4x10-8	0.49(0.38-0.63)	NA	NA	NA	NA

CHR	Chromosome
SNP	Single nucleoitide polymorphism
BP	Biase paor
FCAUK	P value UK
FCOUK	Allele freqency in UK Controles
PUK	Allele freqency in UK CFS Cases
ORUK	Odds ratio with 95% confidence interval in UK Biobank
FCANOR	Allele freqency in Norwegian ME/CFS Cases
FCONOR	Allele freqency in Norwegian Controles
PNOR	P value Norwegian
OR(95%CI)NOR	Odds ratio with 95% confidence interval in Norwegian

Supplementary Table 8. Most associated SNPs from the meta-analysis of the Norwegian and UK Biobank datasets

CHR	BP	SNP	Closest gene	Position	P_{META}	PR_{META}	OR_{META}	ORR_{META}	Q	I	P_{UK}	OR_{UK}	P_{NOR}	OR_{NOR}	Type of SNP in the Norwegian	Type of SNP in the UK Biobank
6	94006318	rs72914217	*EPHA7*	Intronic	1.421x10-6	1.421x10-6	1.94	1.94	0.8585	0	4.7x10-6	1.3(1.15-1.44)	0.12	1.25(0.94-1.68)	Imputed	imputed
6	94005099	rs17448051	*EPHA7*	Intronic	2.625x10-6	2.625x10-6	1.279	1.279	0.9049	0	8.7x10-6	1.3(1.15-1.44)	0.12	1.25(0.94-1.68)	Imputed	imputed
6	94004630	rs17545130	*EPHA7*	Intronic	2.71x10-6	2.71x10-6	1.48	1.48	0.9601	0	1.0x10-5	1.3(1.14-1.42)	0.10	1.26(0.94-1.68)	Imputed	imputed
6	94010593	rs72914220	*EPHA7*	Intronic	3.183x10-6	3.183x10-6	1.98	1.98	0.9011	0	1.0x10-5	1.3(1.14-1.43)	0.12	1.25(0.94-1.68)	Imputed	imputed
17	46293875	rs7221416	*SKAP1*	Intronic	4.627x10-6	4.627x10-6	1.66	1.66	0.663	0	5.06x10-5	1.18(1.09-1.29)	0.02	1.24(1.02-1.51)	Imputed	imputed
17	46261239	rs1316643	*SKAP1*	Intronic	5.445x10-6	5.445x10-6	1.46	1.46	0.6253	0	6.01x10-5	1.18(1.09-1.29)	0.02	1.25(1.02-1.52)	Imputed	imputed
17	46206211	rs5820685	*SKAP1*	Intronic	6.257x10-6	6.257x10-6	1.26	1.26	0.6297	0	7.4x10-5	1.18(1.09-1.29)	0.02	1.25(1.02-1.52)	Imputed	Imputed
17	46238276	rs11079812	*SKAP1*	Intronic	6.809x10-6	6.809x10-6	1.192	1.192	0.6083	0	7.9x10-5	1.18(1.09-1.29)	0.02	1.25(1.02-1.52)	Imputed	Genotyped
17	46231780	rs35371915	*SKAP1*	Intronic	8.154x10-6	8.154x10-6	1.03	1.03	0.5975	0	9.4x10-5	1.18(1.09-1.29)	0.02	1.25(1.02-1.52)	Imputed	imputed
17	46259254	rs12948653	*SKAP1*	Intronic	8.265x10-6	8.265x10-6	1.03	1.03	0.5982	0	9.4x10-5	1.18(1.09-1.29)	0.02	1.25(1.02-1.52)	Imputed	imputed

13	41404706	rs11147812	*TPTE2P5*	Intronic	8.772x10-6	1.468x10-5	0.8587	0.8591	0.3131	1.72	7.7x10-6	0.84(0.78-0.91)	0.36	0.92(0.78-1.03)	Imputed	imputed
17	46223379	rs12449324	*SKAP1*	Intronic	8.811x10-6	8.811x10-6	1.1888	1.1888	0.6215	0	9.9x10-5	1.18(1.09-1.29)	0.02	1.25(1.02-1.50)	Imputed	imputed
17	46243167	rs17619103	*SKAP1*	Intronic	8.972x10-6	8.972x10-6	1.1894	1.1894	0.5928	0	9.7x10-5	1.18(1.09-1.29)	0.02	1.25(1.02-1.50)	Imputed	imputed
22	51105101	rs1858756	*SHANK3*	Intergenic	9.04x10-6	0.003427	1.1617	1.1495	0.2329	29.74	5.4x10-6	1.18(1.1-1.27)	0.46	0.94(0.80-1.1)	Imputed	imputed
17	39300697	rs139894014	*KRTAP4-6/KRTAP4-5*	Intergenic	9.233x10-6	9.233x10-6	1.02	1.02	0.9579	0	1.8x10-5	1.32(1.16-1.51)	0.22	1.34(0.83-2.18)	Imputed	imputed
22	51098443	rs9616900	*SHANK3*	Intergenic	9.431x10-6	0.0004424	1.163	1.1567	0.2732	16.71	7.3x10-6	1.18(1.1-1.28)	0.39	0.93(0.79-1.09)	Imputed	imputed
17	46219650	rs35999311	*SKAP1*	Intronic	9.596x10-6	9.596x10-6	1.1879	1.1879	0.616	0	0.0001	1.17(1.08-1.28)	0.02	1.25(1.02-1.50)	imputed	imputed
22	51089624	rs11704245	*SHANK3*	Intergenic	9.68x10-6	0.003541	1.1661	1.1534	0.2285	31.04	6.3x10-6	1.18(1.1-1.28)	0.43	1.06(0.90-1.25)	imputed	imputed
17	46233490	rs11079811	*SKAP1*	Intronic	9.752x10-6	9.752x10-6	1.1886	1.1886	0.5868	0	0.0001	1.17(1.08-1.28)	0.02	1.25(1.02-1.50)	imputed	imputed
19	55897327	rs17700376	RPL28	Exonic	1.008x10-5	1.008x10-5	1.1675	1.1675	0.9787	0	5.2x10-5	1.16(1.08-1.25)	0.07	1.17(0.98-1.39)	imputed	Genotyped

CHR	Chromosome
BP	Single nucleoitide polymorphism
SNP	Base pair position
Closest gene	Closest annotated gene
Position	Genomic region
PMETA	P value in the fixed effect size model
PRMETA	P value in the random effect size model
ORMETA	Odds ratio in the fixed effect size model
ORRMETA	Odds ratio in the random effect size model

Q	P value for Cochrane's Q statistic
I	I^2 heterogeneity index (0-100)
PUK	P value in the UK Biobank
ORUK	Odds ratio in the UK Biobank
PNOR	P value in the Norwegian
ORNOR	Odds ratio in the Norwegian
Type of SNP in the Norwegian	Method of the obtained genotype
Type of SNP in the UK Biobank	Method of the obtained genotype

Supplementary table 9. Association analysis of candidate SNPs in key immune genes previously reported

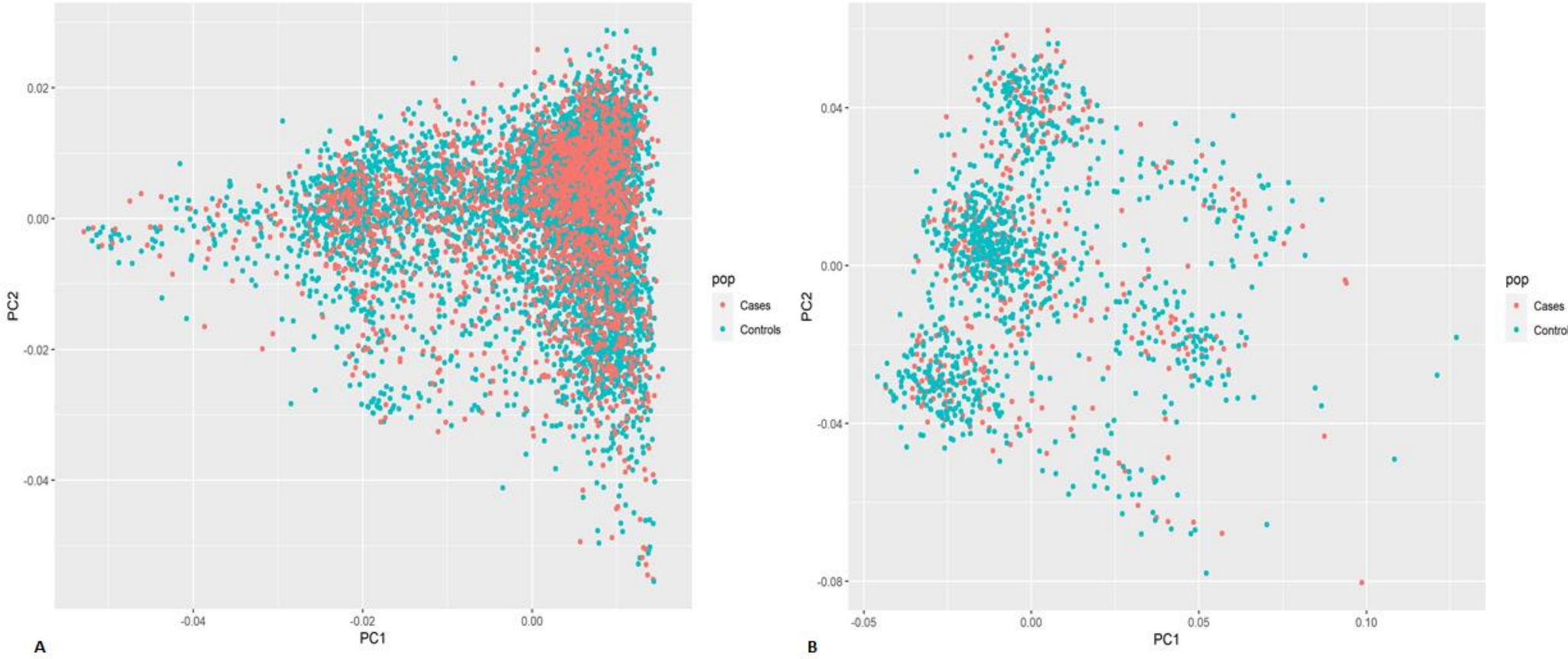
				Data published in Steiner et al, Front Immunol, 2020				**Norwegian cohort**								**UK biobank**			
CHR	SNP	BP	Gene	Af_{cases}	$AF_{controls}$	P	OR	$AF_{cases\ all}$	$AF_{controls\ all}$	P_{all}	OR_{all}	$AF_{cases\ infection}$	$AF_{controls\ infection}$	$P_{infection}$	$OR_{infection}$	AF_{Cases}	$AF_{controls}$	P	OR
1	rs2476601	114377568	*PTPN22*	0.13	0.08	0.016	1.64	0.119	0.121	0.91	0.99	0.118	0.121	0.84	0.97	0.106	0.103	0.68	01.03
2	rs3087243	204738919	*CTLA4*	0.66	0.56	0.001	1.53	0.374	0.399	0.23	0.90	0.387	0.399	0.62	0.95	0.450	0.455	0.66	0.98
6	rs1799724	31542482	*TNF*	0.11	0.13	0.204	0.84	0.079	0.063	0.13	2.32	0.074	0.063	0.33	1.20	0.076	0.074	0.68	01.03
6	rs1800629	31543031	*TNF*	0.14	0.16	0.275	0.89	0.165	0.180	0.34	0.90	0.165	0.180	0.40	0.90	0.190	0.189	0.96	1.00
7	rs3807306	128580680	*IRF5*	0.49	0.51	0.318	0.94	0.492	0.489	0.88	1.01	0.498	0.489	0.68	01.04	0.496	0.502	0.50	0.98

CHR	Chromosome
SNP	SNP
BP	Base pair position
Gene	Gene name
Afcases	Allele frequency of cases in Steiner et.al
AFcontrols	Allele frequency of controls in Steiner et.al
P	P value in Steiner et.al
OR	Odds ratio in Steiner et.al
AFcases all	Allele frequency of cases in our Norwegian cohort of N=427
AFcontrols all	Allele frequency of controls n our Norwegian cohort of N=427

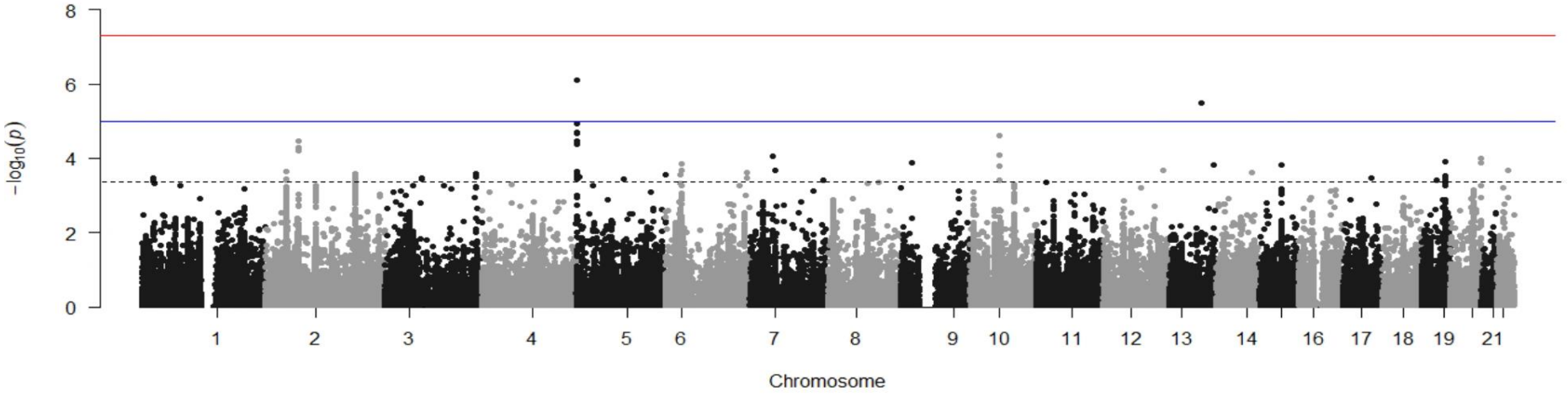
Pall	P value in our Norwegian cohort of N=427
ORall	Odds ratio in our Norwegian cohort of N=427
AFcases infection	Allele frequency of cases in our Norwegian cohort of N=310 patients reporting disease after infectious onset
AFcontrols infection	Allele frequency of controls in our Norwegian cohort N=310 patients reporting disease after infectious onset
Pinfection	P value in our Norwegian cohort of N=310 patients reporting disease after infectious onset
ORinfection	Odds ratio in our Norwegian cohort of N=310 patients reporting disease after infectious onset
AFCases	Allele frequency of cases in the UK Biobank cohort
AFcontrols	Allele frequency of controls in the UK Biobank cohort
P	P valuein the UK Biobank cohort
OR	Odd ration in the UK Biobank cohort

Supplementary Figures

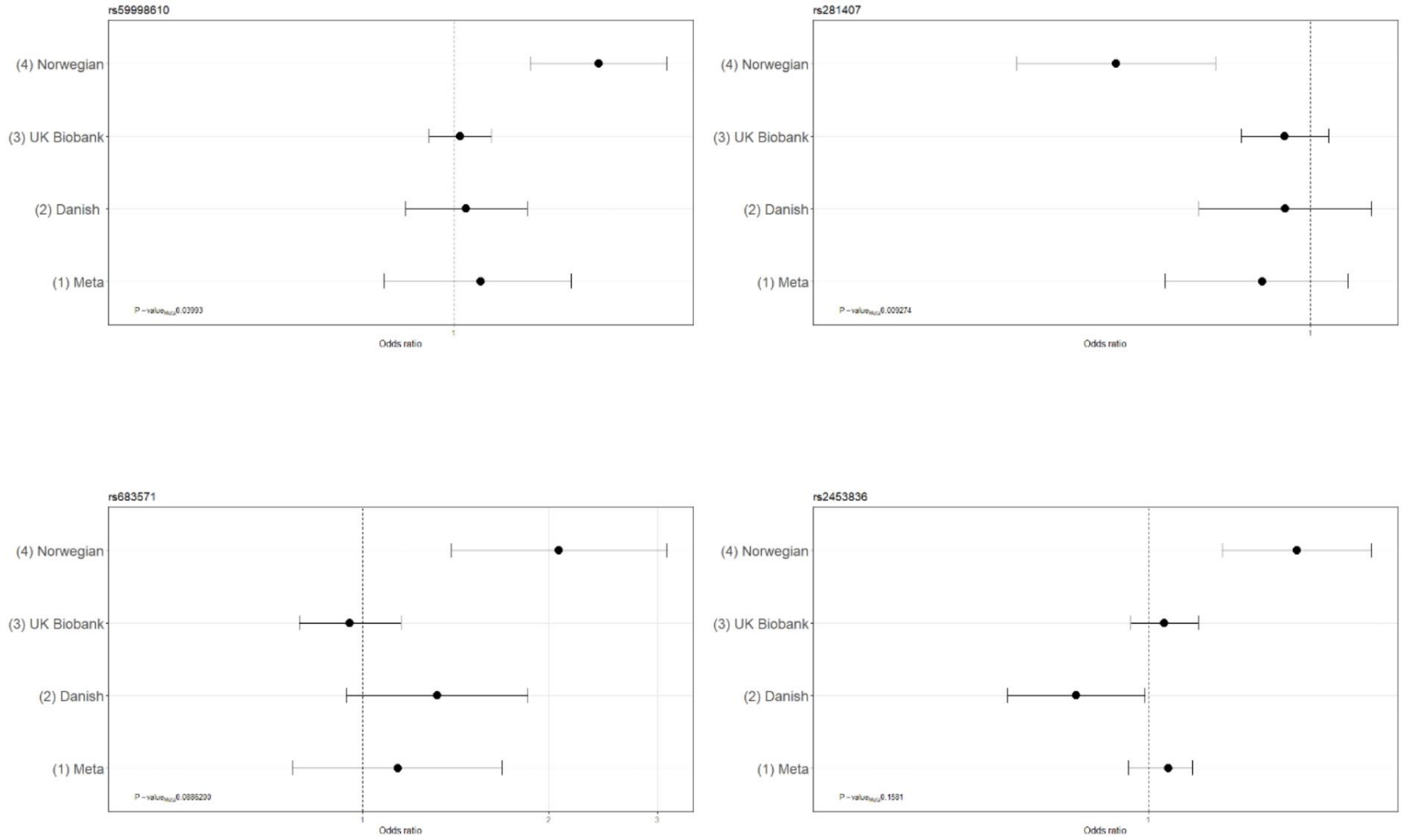
Supplementary FigureS1. A) Principal component plot of the UK Biobank cohort and B) Norwegian discovery cohort

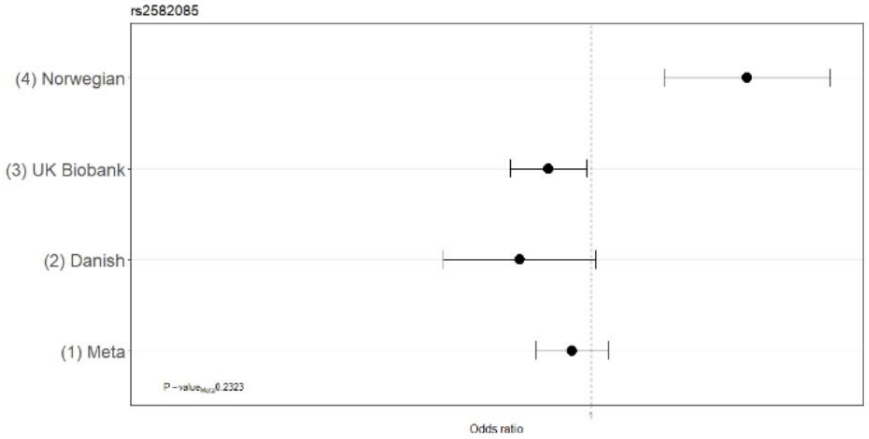


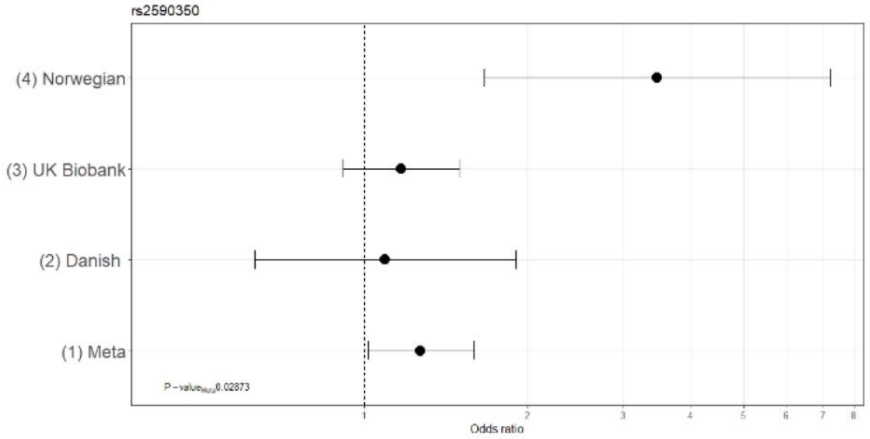
Supplementary figure S2. Manhattan plot of SNP associations across the 22 autosomes for the directly genotyped SNPs (105,902) in 427 Norwegian myalgic encephalomyelitis /chronic fatigue syndrome (ME/CFS) patients and 972 healthy controls. The statistical significance of the association analysis as -log10 of the P-value (y-axis), is plotted against the chromosomal position of each chromosome in base pairs (bp, x-axis). The red, horizontal line represents a genome wide significance threshold of P=5 x10-8, the blue line represents the suggestive significance level of P=1x10-5 and the dotted grey line represents the inclusion threshold for replication (P= 0.0003). Positions are according to National Center for Biotechnology Information's build 37 (hg19).

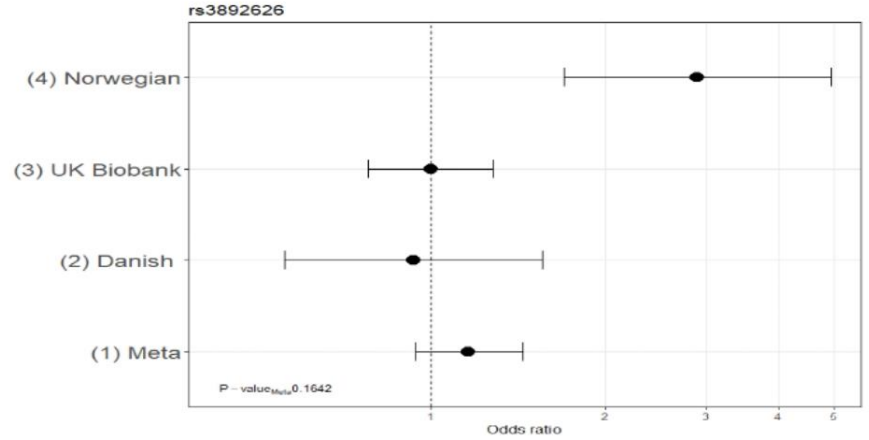


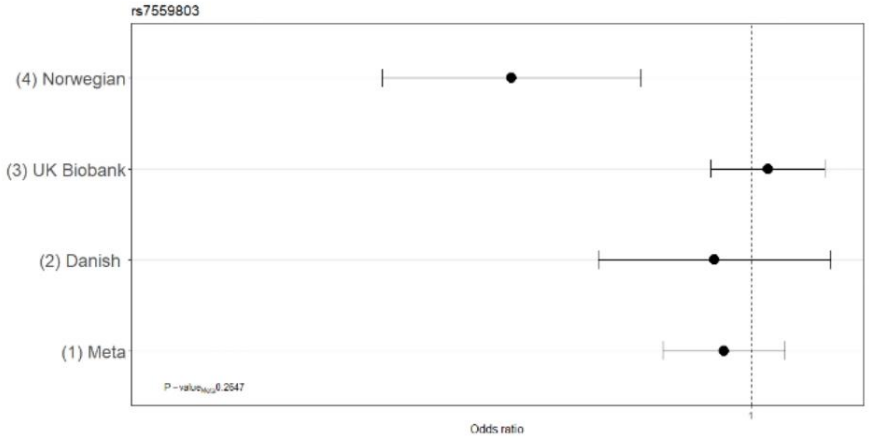
Supplementary figure S3. Odds ratios plots for the SNPs not reaching significance in the combined analysis of the three datasets.

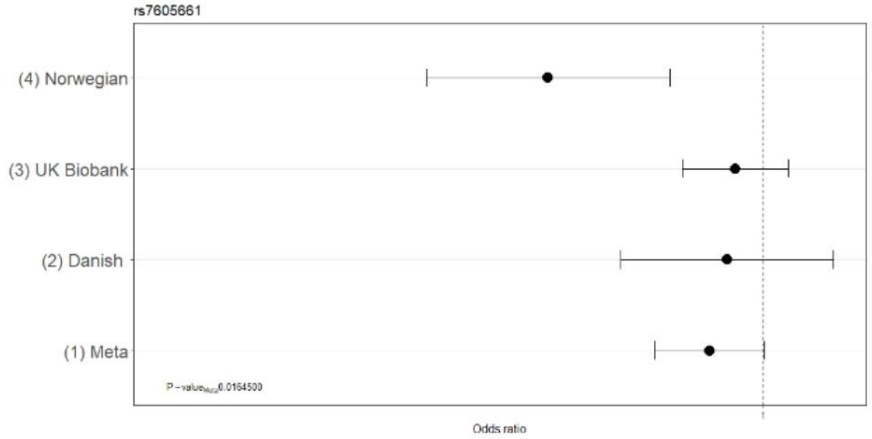


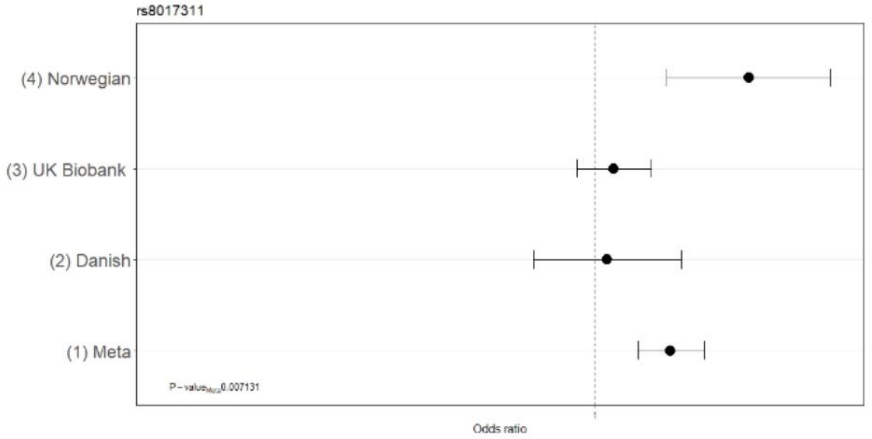

rs2582085
(4) Norwegian
(3) UK Biobank
(2) Danish
(1) Meta
P-value_MU2 0.2323
Odds ratio

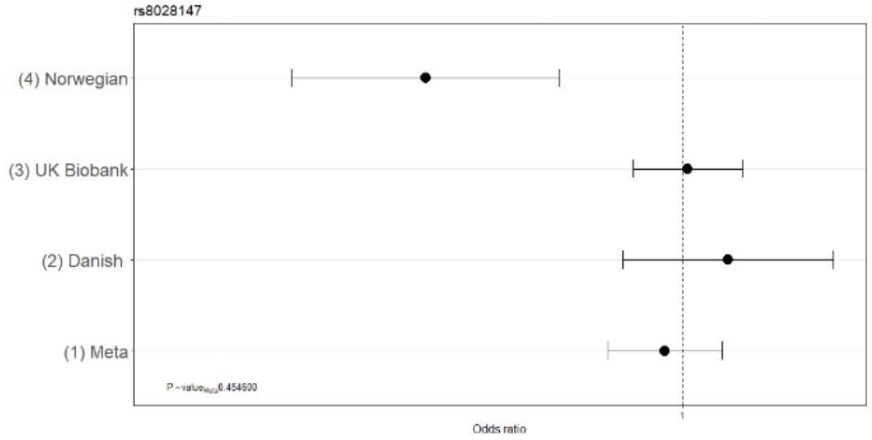

rs2590350
(4) Norwegian
(3) UK Biobank
(2) Danish
(1) Meta
P-value_MU2 0.02873
Odds ratio

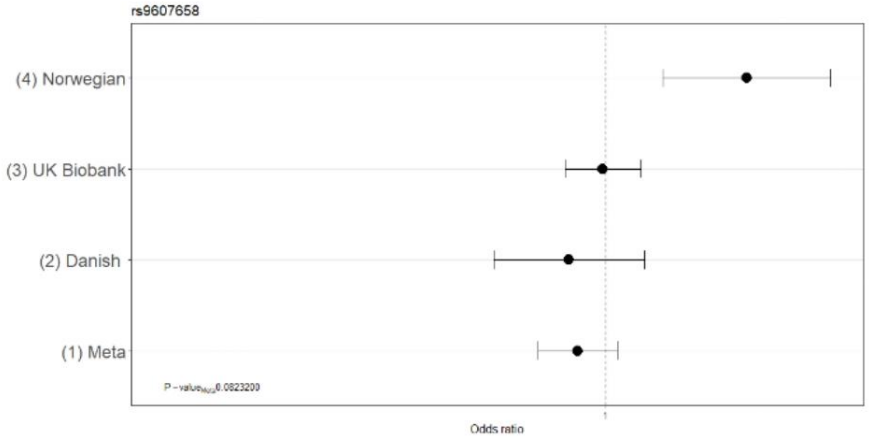

rs3892626
(4) Norwegian
(3) UK Biobank
(2) Danish
(1) Meta
P-value_Meta 0.1642
Odds ratio

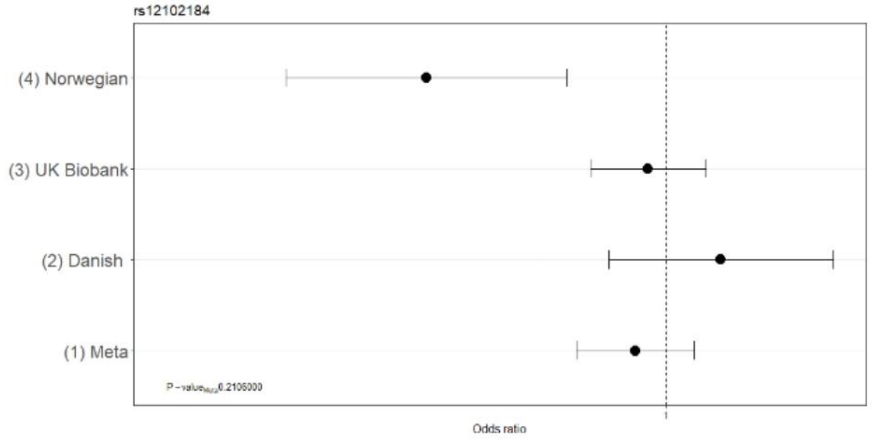

rs7559803
(4) Norwegian
(3) UK Biobank
(2) Danish
(1) Meta
P-value_MU2 0.2647
Odds ratio

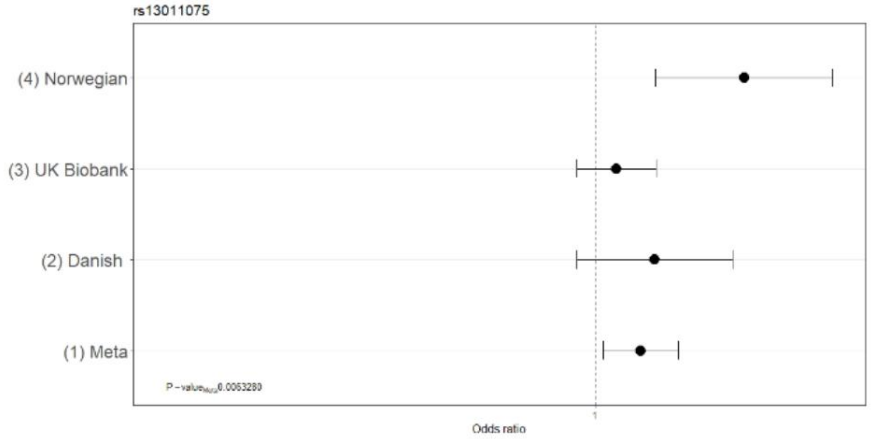

rs7605661
(4) Norwegian
(3) UK Biobank
(2) Danish
(1) Meta
P – value_MVZ 0.0164500
Odds ratio

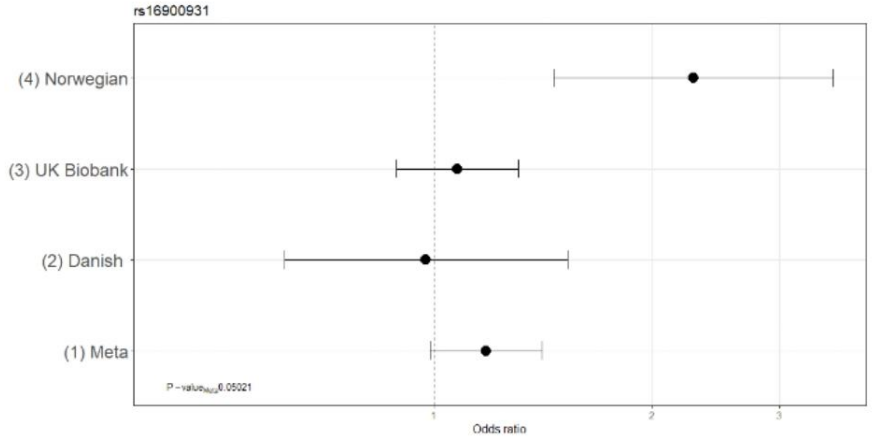

rs8017311
(4) Norwegian
(3) UK Biobank
(2) Danish
(1) Meta
P – value_MVZ 0.007131
Odds ratio

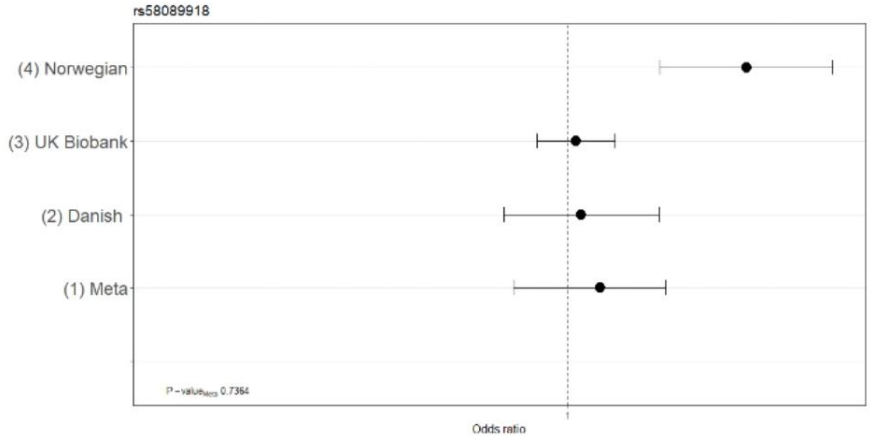

rs8028147
(4) Norwegian
(3) UK Biobank
(2) Danish
(1) Meta
P – value_MVZ 0.454500
Odds ratio


rs9607658
(4) Norwegian
(3) UK Biobank
(2) Danish
(1) Meta
P – value_MVZ 0.0823200
Odds ratio

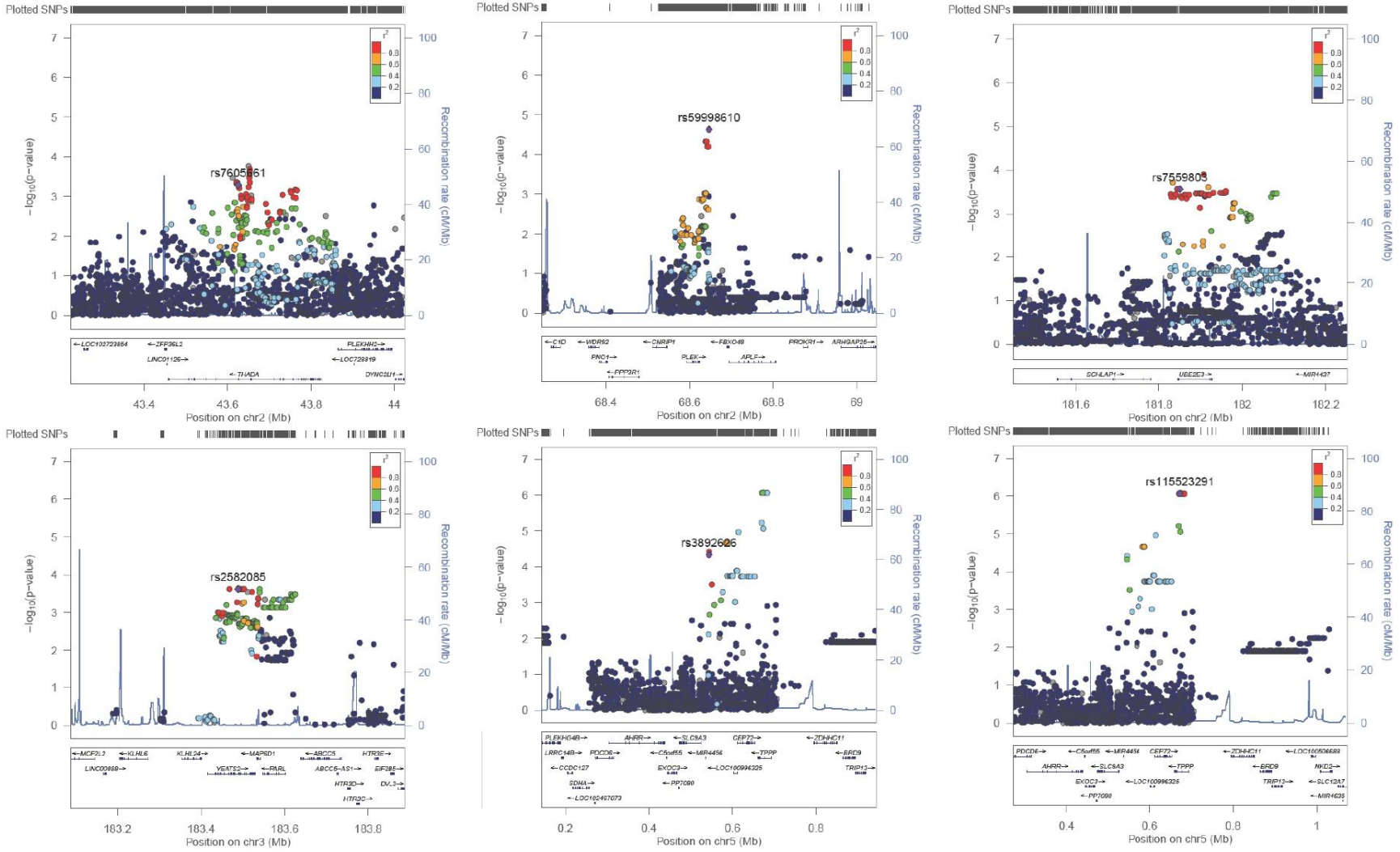

rs12102184
(4) Norwegian
(3) UK Biobank
(2) Danish
(1) Meta
Odds ratio

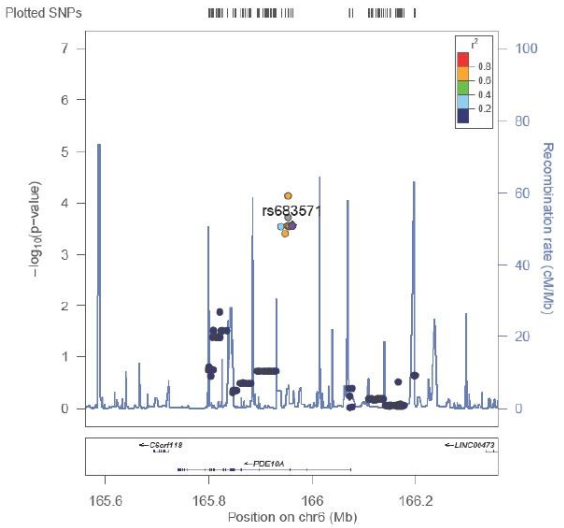

rs13011075
(4) Norwegian
(3) UK Biobank
(2) Danish
(1) Meta
Odds ratio

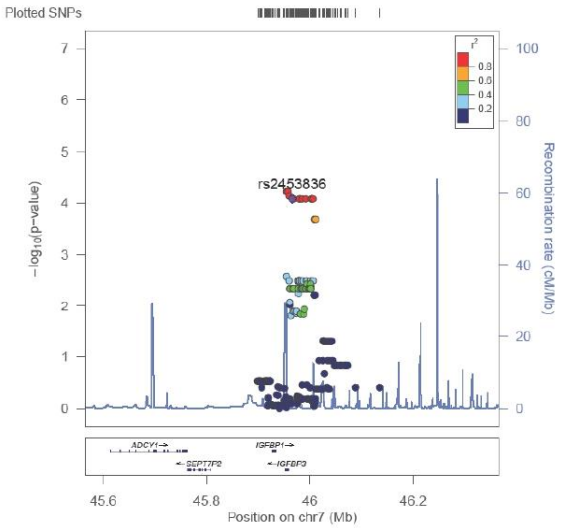

rs16900931
(4) Norwegian
(3) UK Biobank
(2) Danish
(1) Meta
Odds ratio

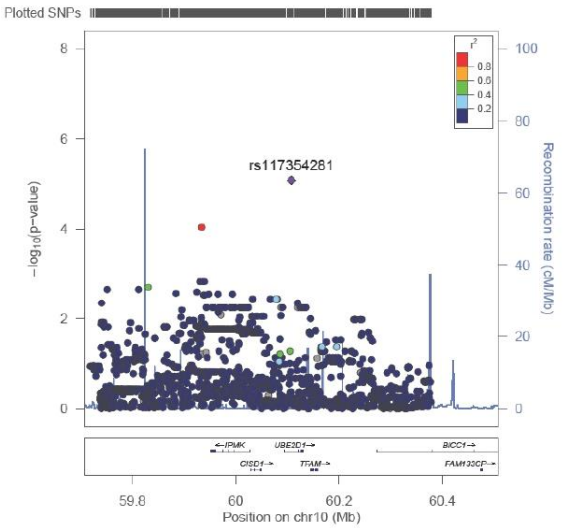

rs58089918
(4) Norwegian
(3) UK Biobank
(2) Danish
(1) Meta
Odds ratio

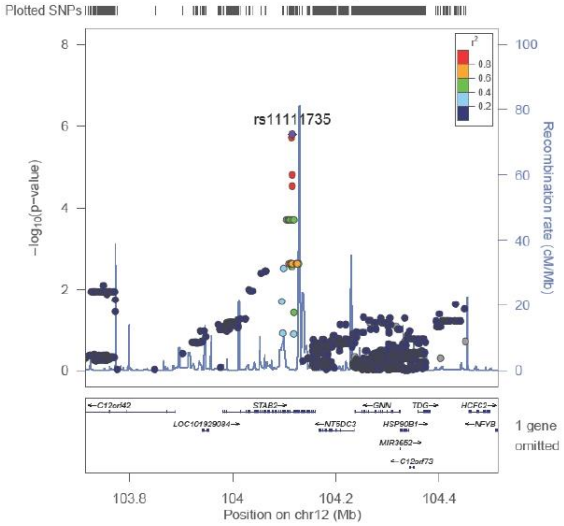
Supplementary figure S4. Plots for the Norwegian dataset for the 20 regions selected for replication including imputed genotypes and showing their LD (r2) with the top SNP within the different regions. The significance of the association signals, -log10P-value (y-axis), are plotted against the positions on the respective chromosome shown in base pairs (bp, x-axis). The colour of the dots indicates r2 value.

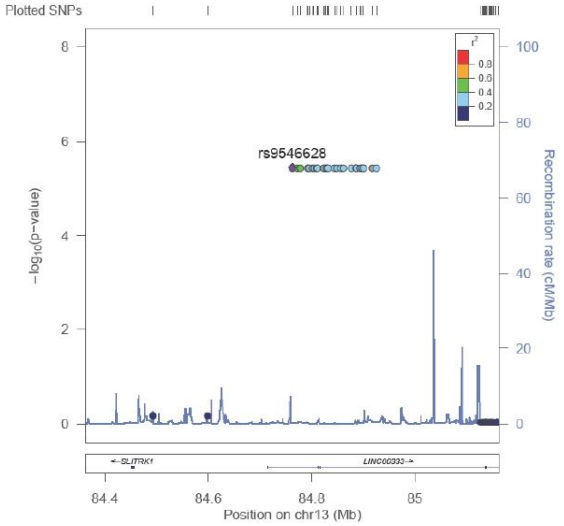


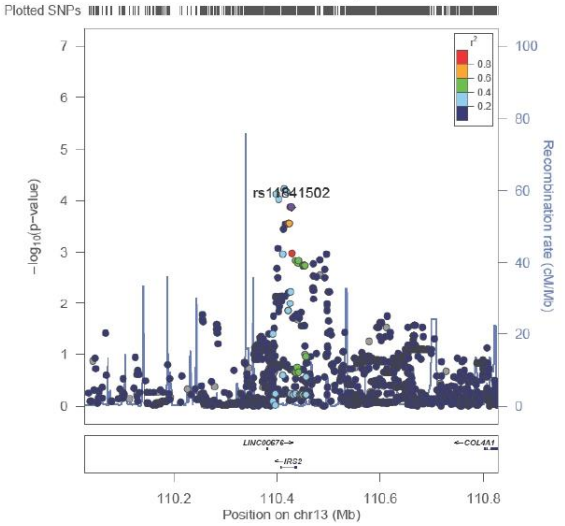

Plotted SNPs
rs683571
−log10(p−value)
Recombination rate (cM/Mb)
Position on chr6 (Mb)

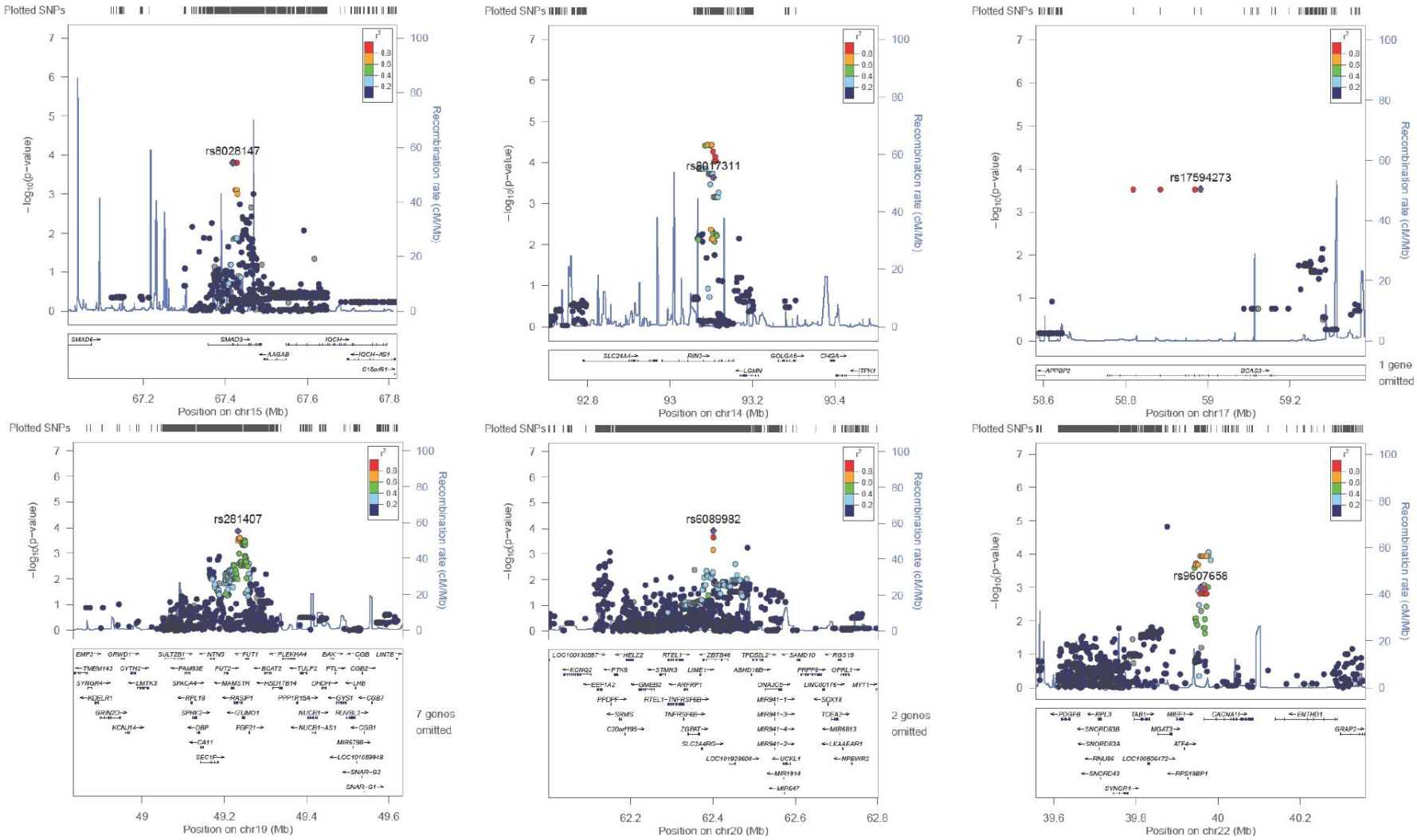

Plotted SNPs
rs2453836
−log10(p−value)
Recombination rate (cM/Mb)
Position on chr7 (Mb)


Plotted SNPs
rs117354281
−log10(p−value)
Recombination rate (cM/Mb)
Position on chr10 (Mb)

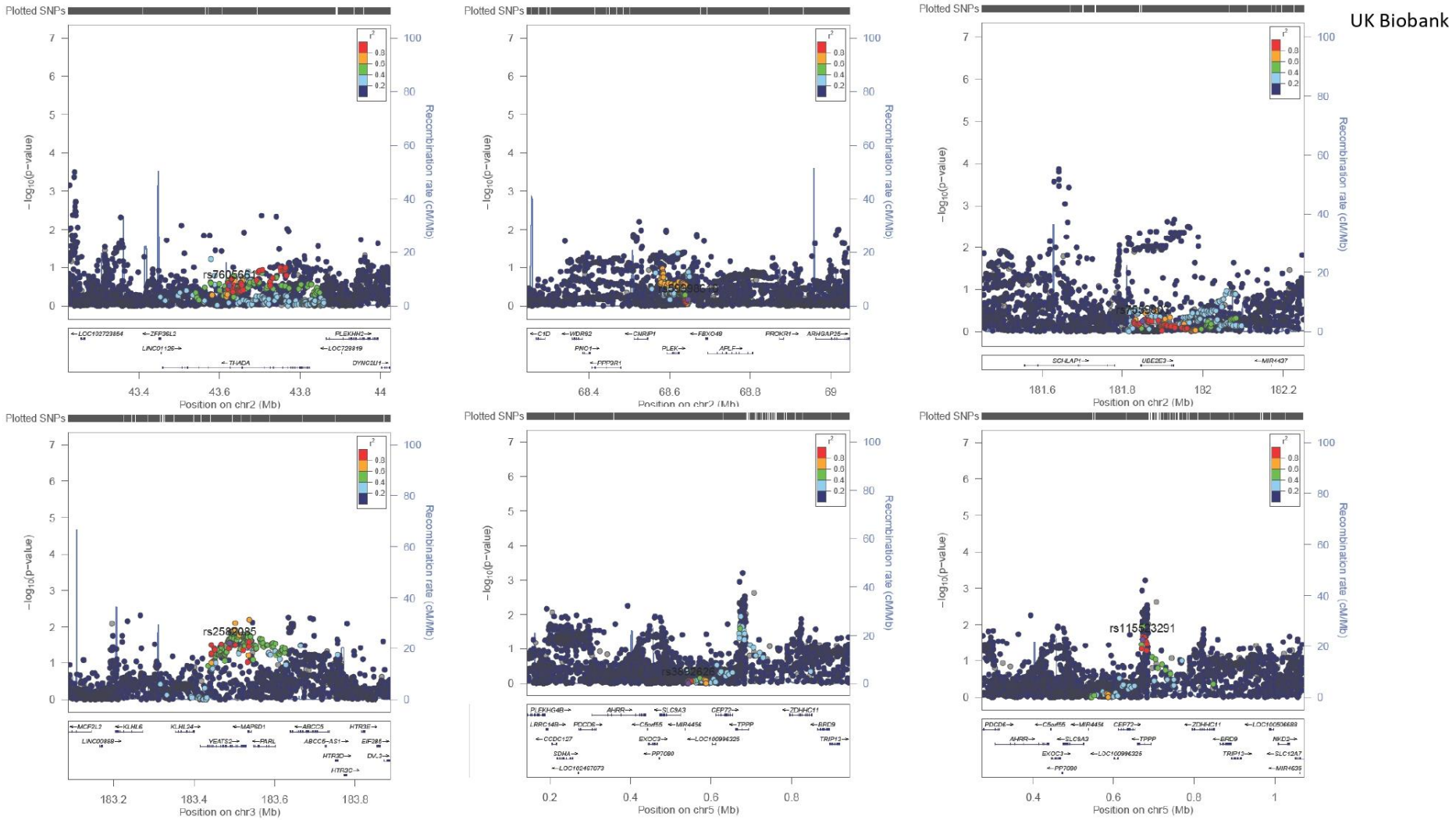

Plotted SNPs
rs11111735
−log10(p−value)
Recombination rate (cM/Mb)
1 gene omitted
Position on chr12 (Mb)


Plotted SNPs
rs9546628
−log10(p−value)
Recombination rate (cM/Mb)
Position on chr13 (Mb)

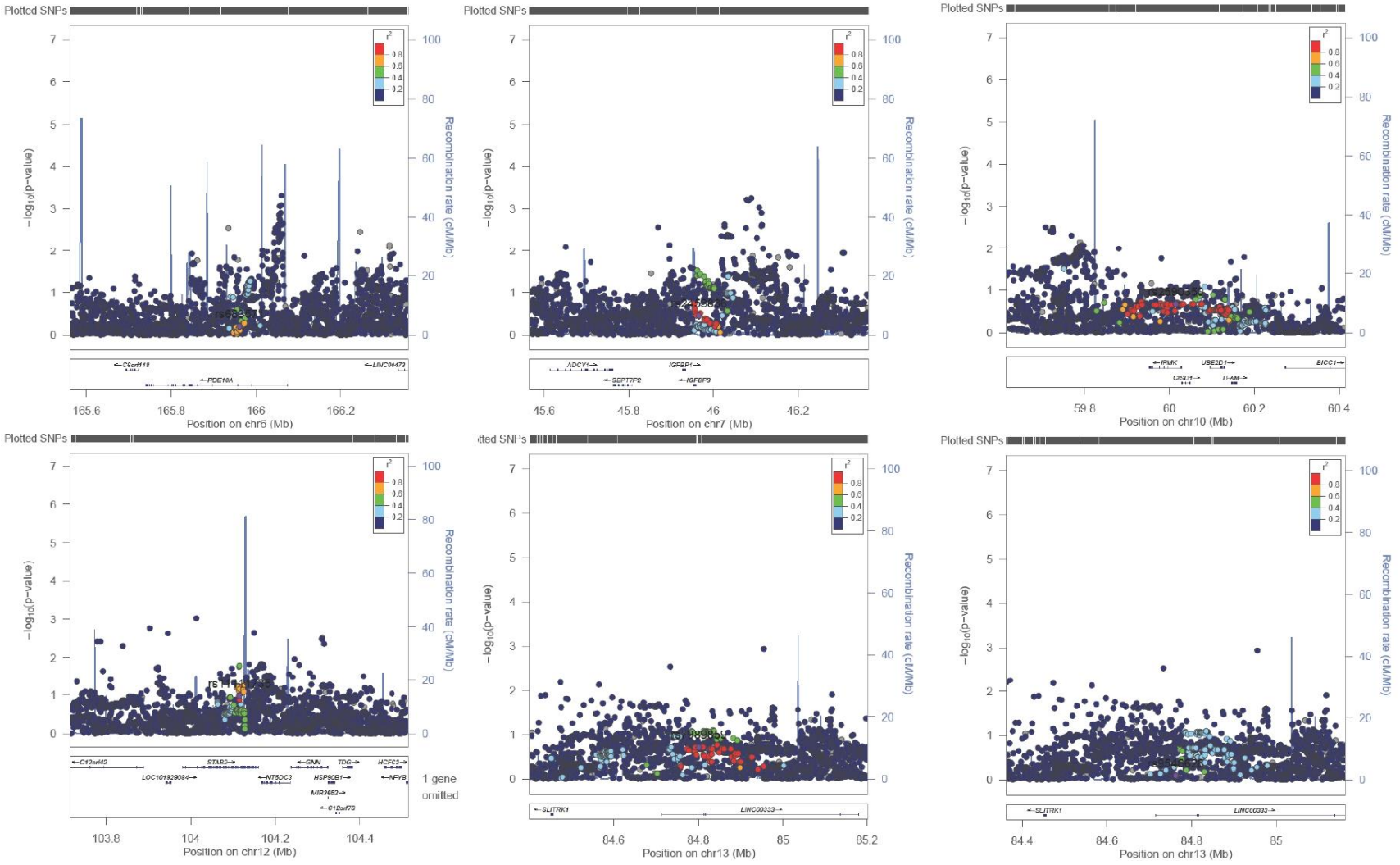

Plotted SNPs
rs10841502
−log10(p−value)
Recombination rate (cM/Mb)
Position on chr13 (Mb)

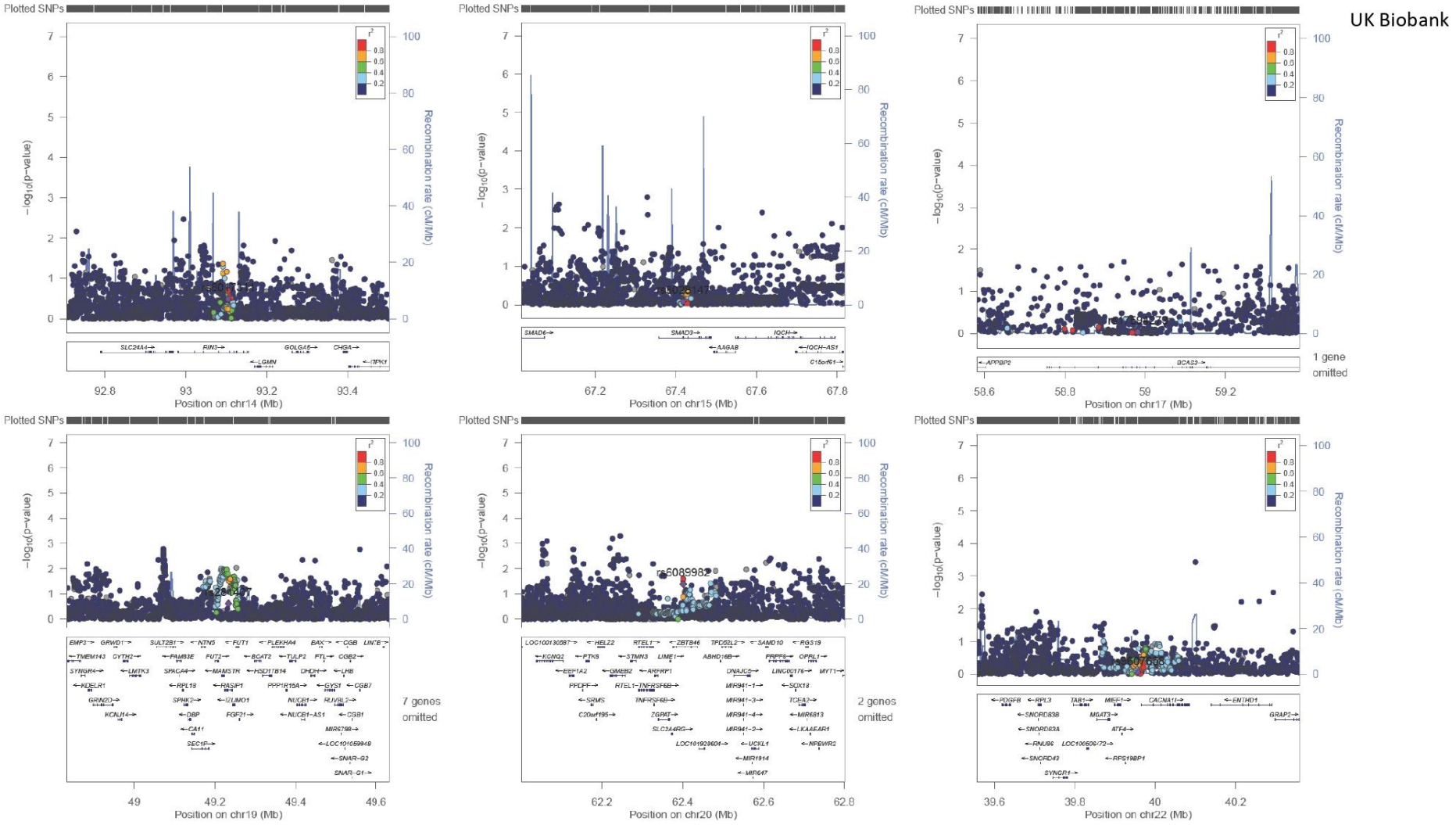

Plotted SNPs
rs8028147
rs6017311
rs17594273
rs281407
rs6089982
rs9607658
−log10(p−value)
Recombination rate (cM/Mb)
Position on chr15 (Mb)
Position on chr14 (Mb)
Position on chr17 (Mb)
Position on chr19 (Mb)
Position on chr20 (Mb)
Position on chr22 (Mb)
1 gene omitted
7 genes omitted
2 genes omitted

Supplementary figure S5. Plots for the UK biobank cohort for the 20 regions selected for replication including imputed genotypes and showing their LD (r2) with the top SNP within the different regions. The significance of the association signals, -log10P-value (y-axis), are plotted against the positions on the respective chromosome shown in base pairs (bp, x-axis). The colour of the dots indicates r2 value.

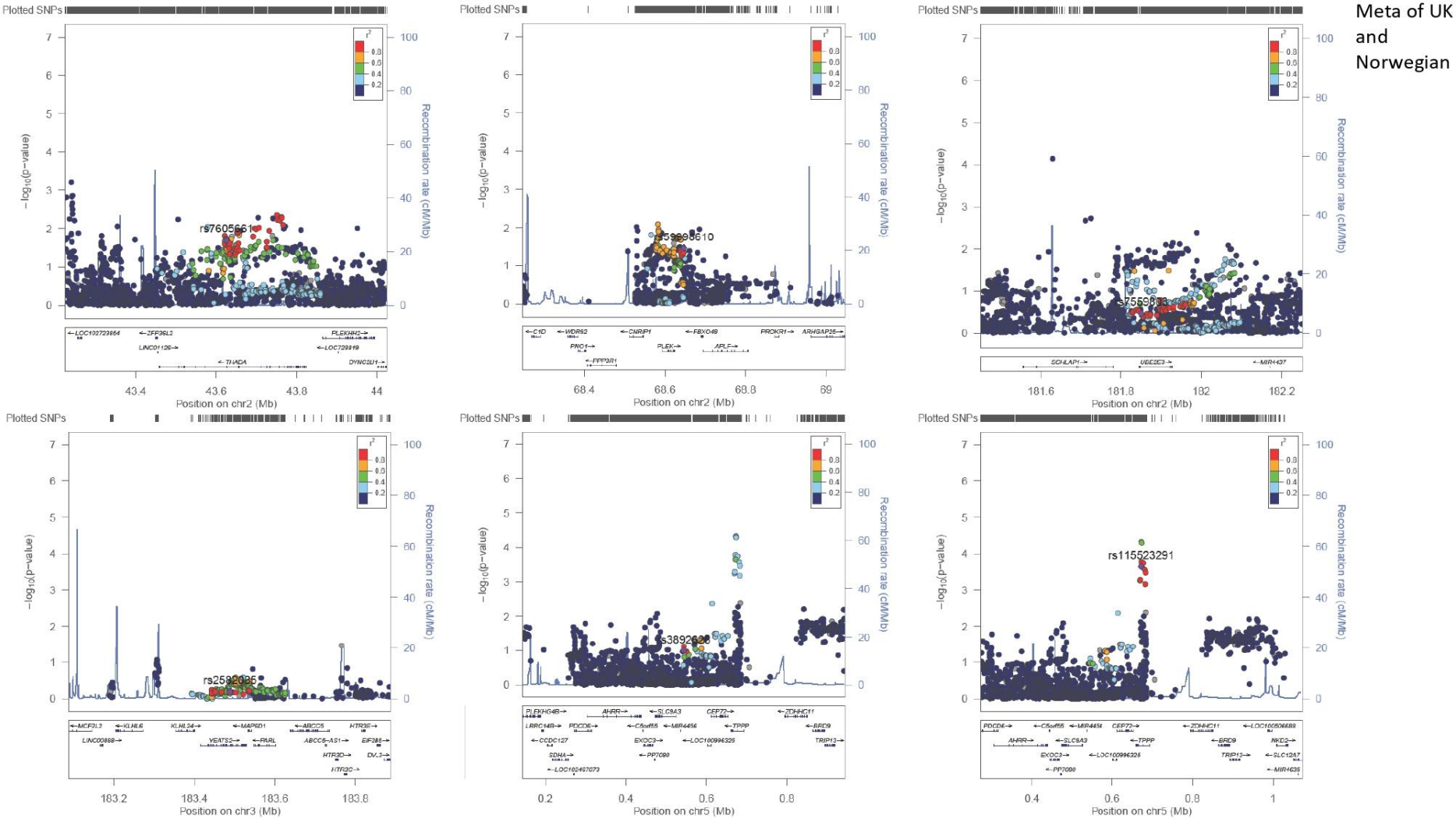


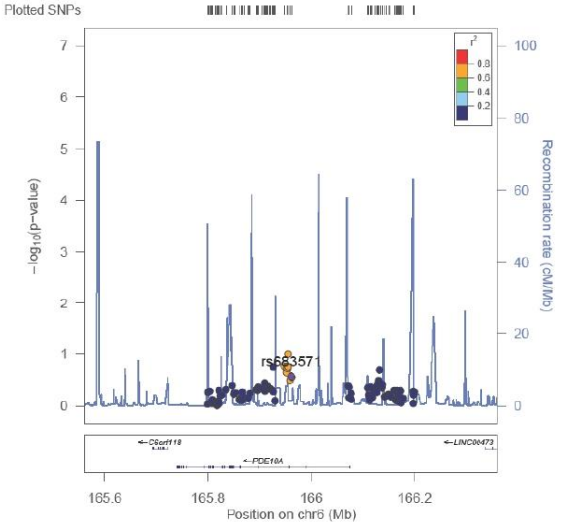
UK Biobank

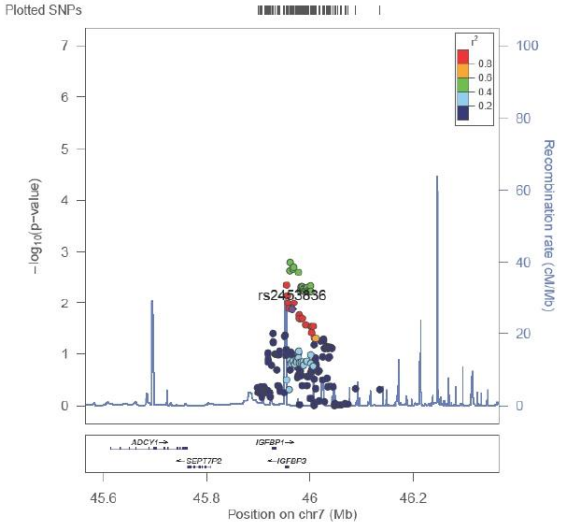

Plotted SNPs
−log10(p-value)
Recombination rate (cM/Mb)
Position on chr6 (Mb)
Position on chr7 (Mb)
Position on chr10 (Mb)
Position on chr12 (Mb)
Position on chr13 (Mb)
Position on chr13 (Mb)
1 gene omitted

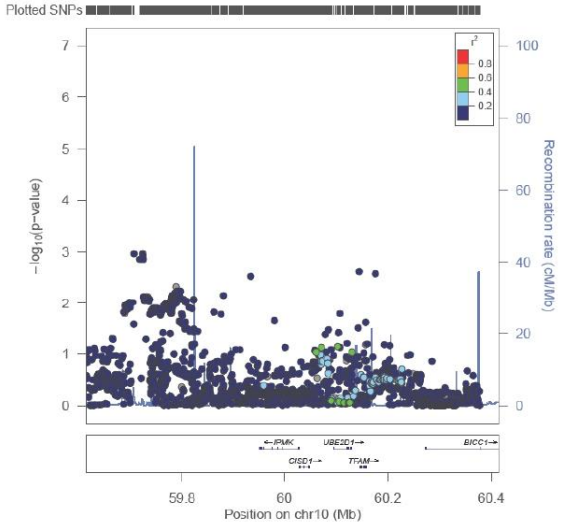

UK Biobank
Plotted SNPs
−log10(p-value)
Recombination rate (cM/Mb)
Position on chr14 (Mb)
Position on chr15 (Mb)
Position on chr17 (Mb)
1 gene omitted
Position on chr19 (Mb)
7 genes omitted
Position on chr20 (Mb)
2 genes omitted
Position on chr22 (Mb)

Supplementary figure S6. Plots based on the combined meta-analysis of the UK biobank and Norwegian cohorts for the 20 regions selected for replication, including imputed genotypes and showing their LD (r^2) with the top SNP within the different regions.
The significance of the association signals, -log10P-value (y-axis), are plotted against the positions on the respective chromosome shown in base pairs (bp, x-axis). The colour of the dots indicates r^2 value.

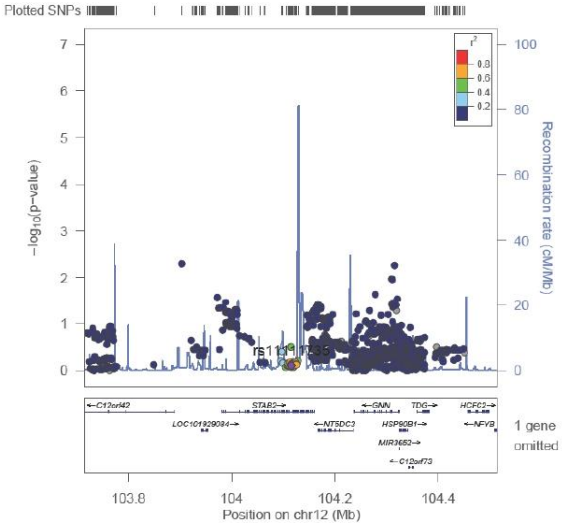


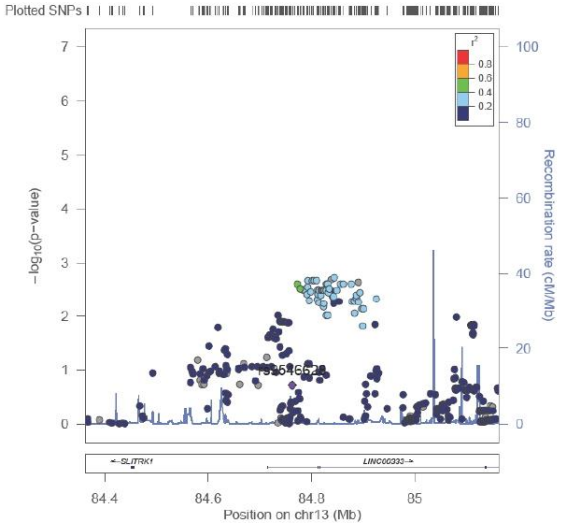


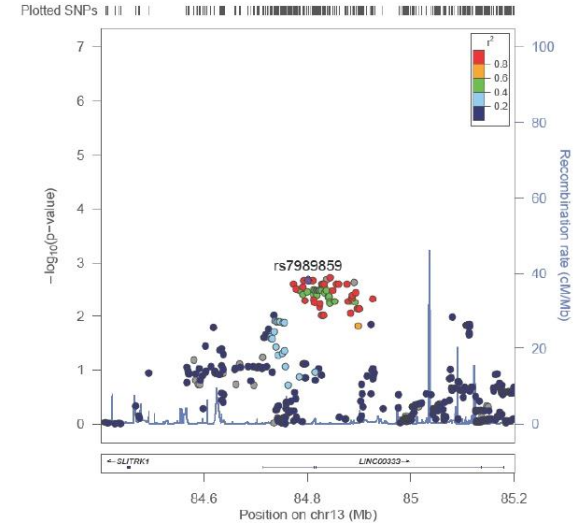


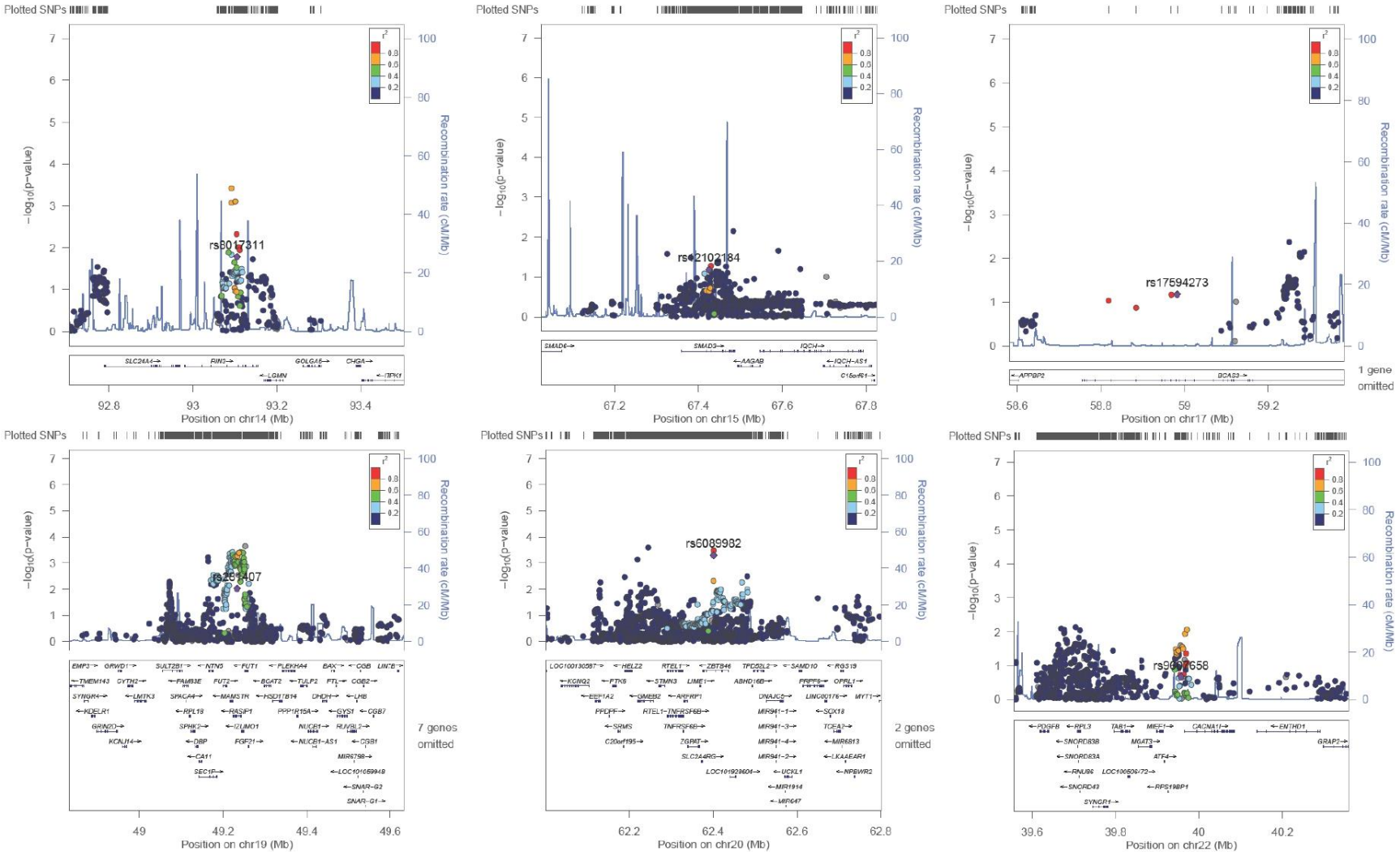


Meta of UK and Norwegian



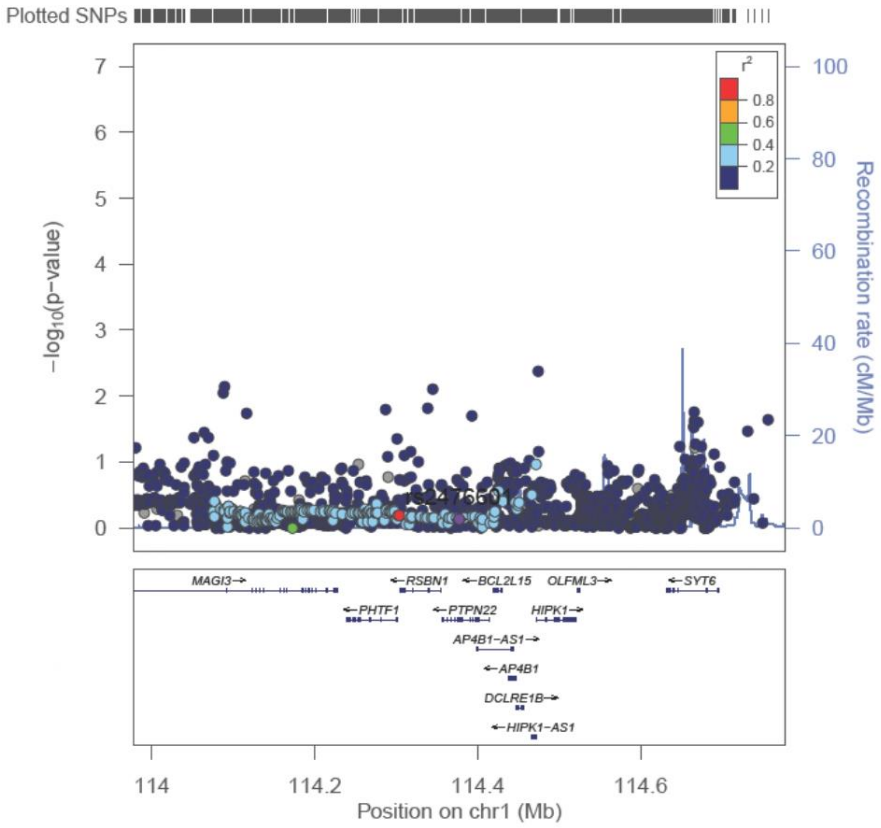


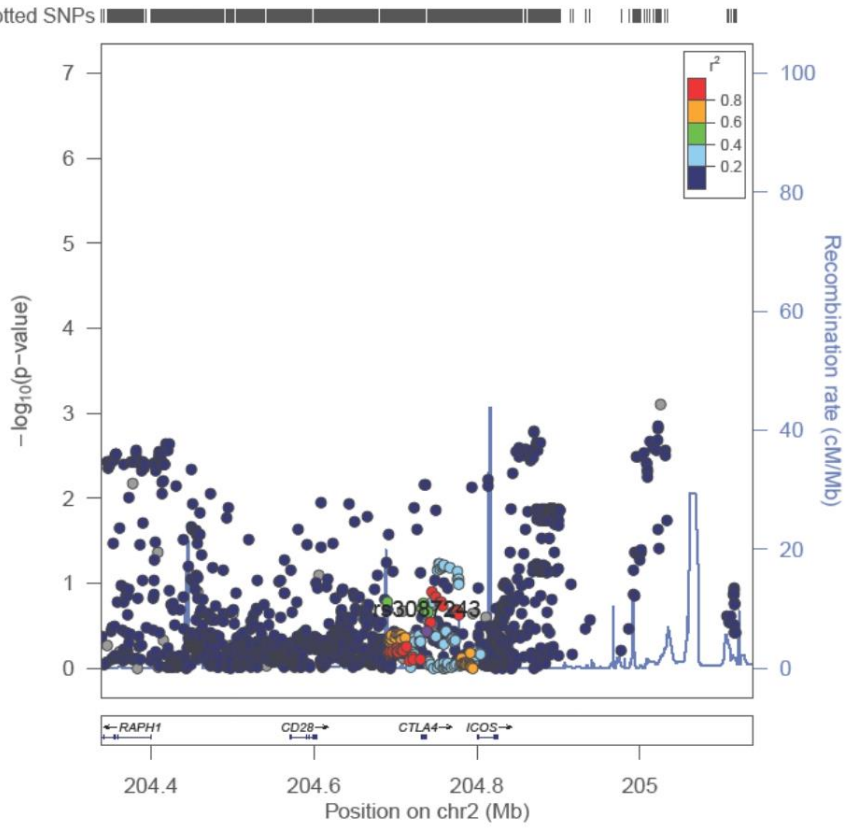




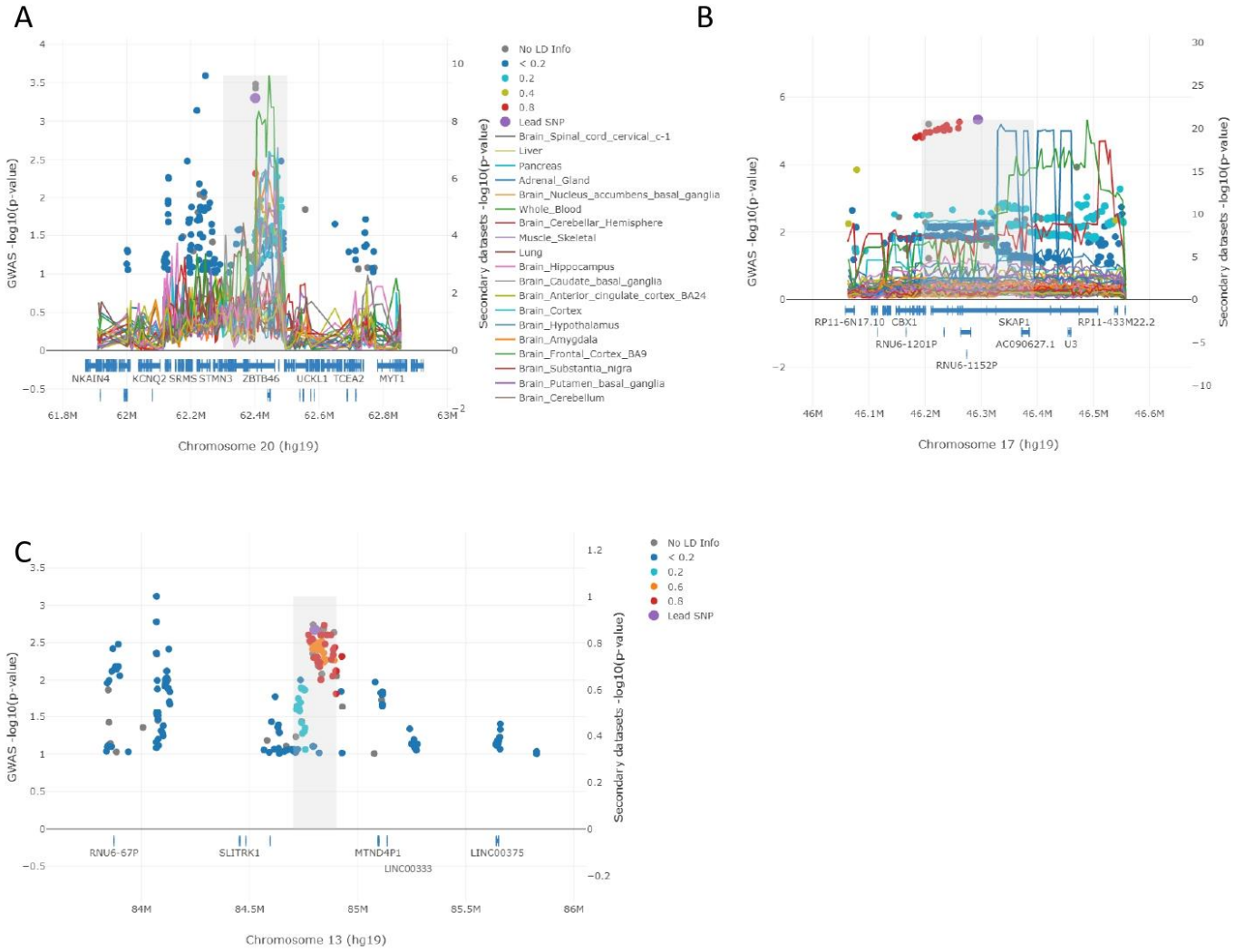
Meta of UK and Norwegian

Supplementary figure S7. Plots based on the combined analyses of the UK biobank the Norwegian cohorts for the *PTPN22* and *CTLA4* regions including imputed genotypes and showing their LD (r2) with the SNPs identified as associated with ME/CFS by Steiner et al.. The significance of the association signals, -log10P-value (y-axis), are plotted against the positions on the respective chromosome shown in base pairs (bp, x-axis). The colour of the dots indicates r2 value.

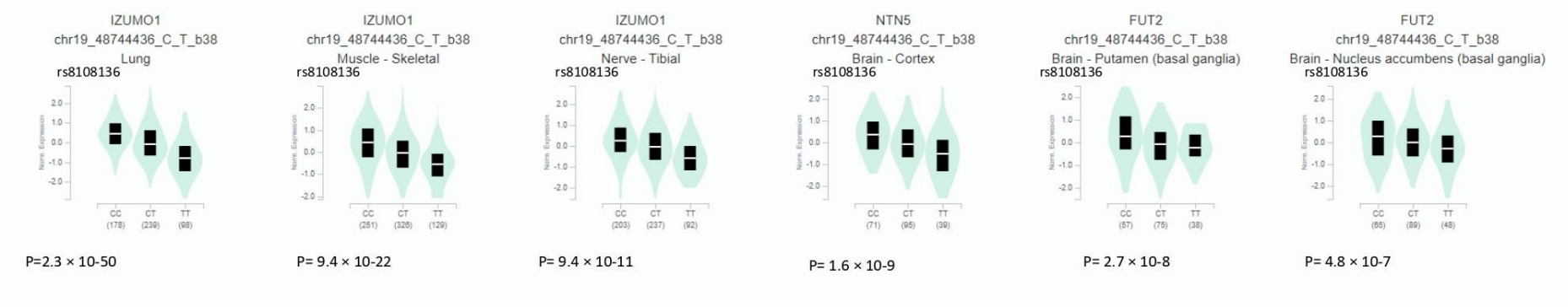




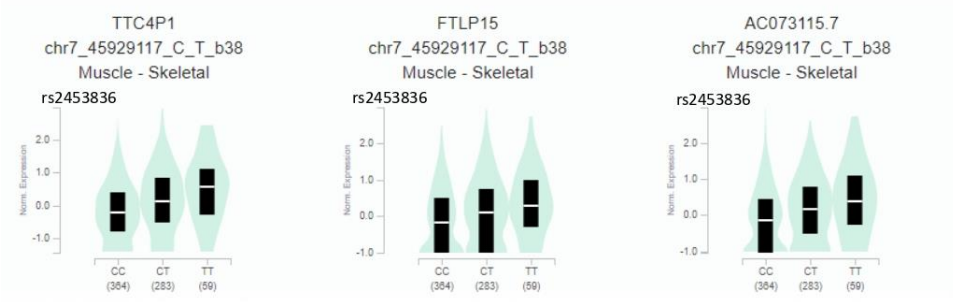
Supplementary figure S8. Colocalization plot generated in LocusFocus for the combined analysis of the UK biobank and Norwegian cohorts showing the degree of pairwise LD with the lead markers (rs2777943, rs368711309, rs7221416). The x-axis represents the chromosomal position in base pairs while the two y-axes represent the significance of the SNP and eQTL associations respectively.



Supplementary figure S9. Expression quantitative trait loci (eQTL) plots for genotypes of tag or top SNPs from the six putative ME/CFS associated regions for genes showing significantly differential expression in relevant tissues. Obtained from GTEx.

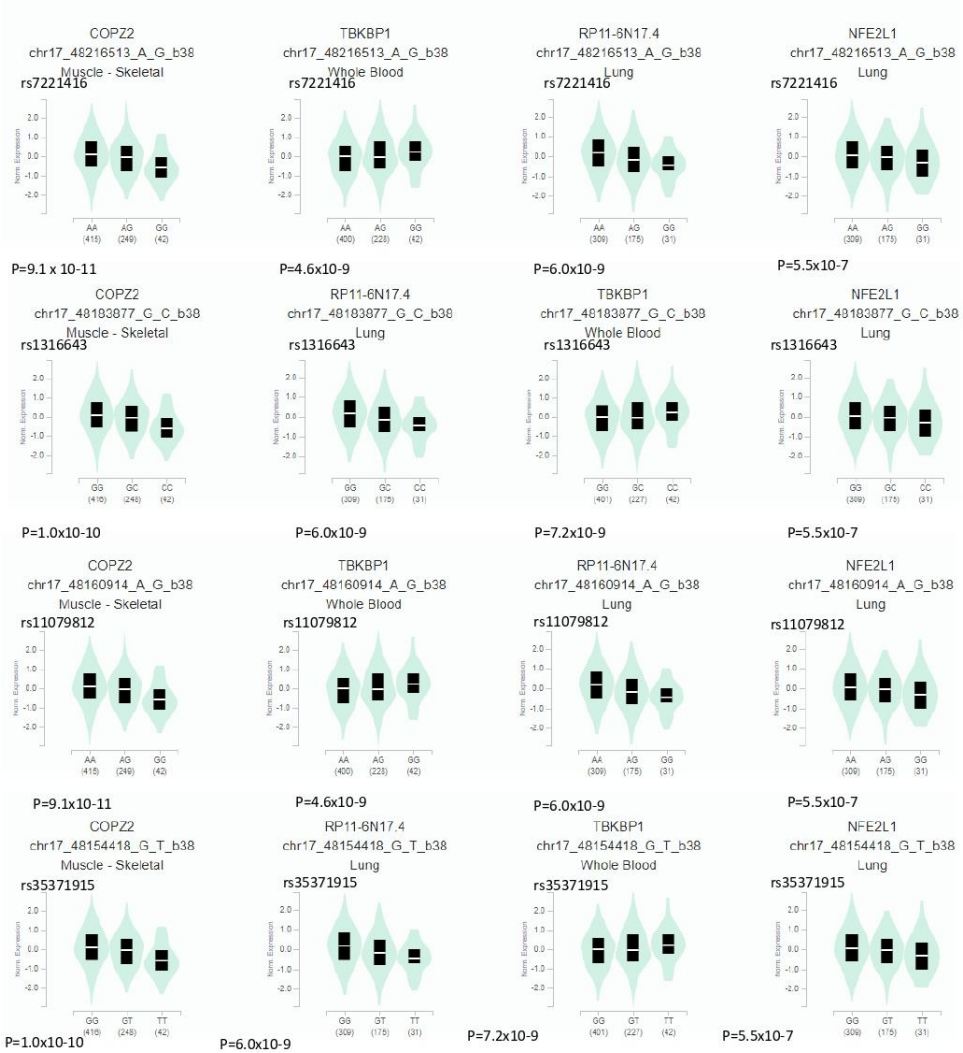


P=2.3 × 10-50 P= 9.4 × 10-22 P= 9.4 × 10-11 P= 1.6 × 10-9 P= 2.7 × 10-8 P= 4.8 × 10-7

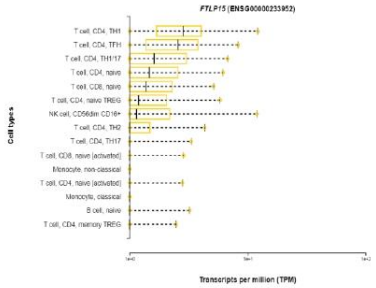


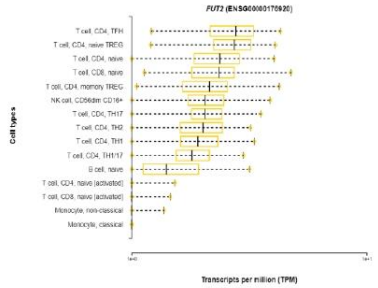
P= 1.1 × 10-15 P= 2.3 × 10-10 P= 1.5× 10-8

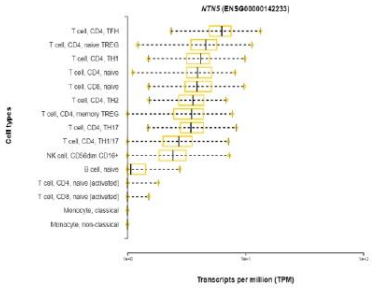
Supplementary figure S10. Expression quantitative trait loci (eQTL) plots for SKAP1 SNPs showing significantly differential expression in relevant tissues. Obtained from GTEx.

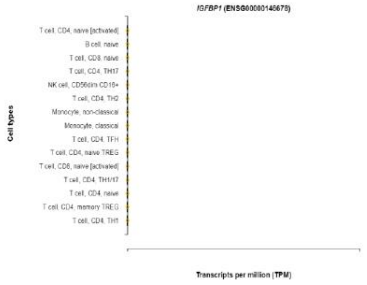


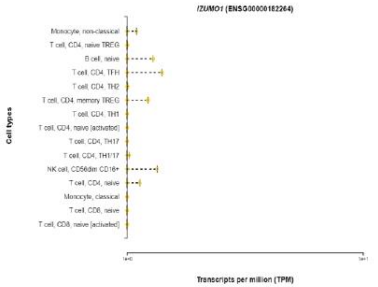
Supplementary figure S11. Expression level plots of the putative ME/CFS associated genes (FTLP15,FUT2, NTN5, IGFBP1, IZUMO1, LINC00333, MAMSTR, TPPP, TTC4P1 and ZBTB46) in immune cells. Obtained via DICE (Database of Immune Cell Expression, Expression quantitative trait loci (eQTLs) and Epigenomics) database.

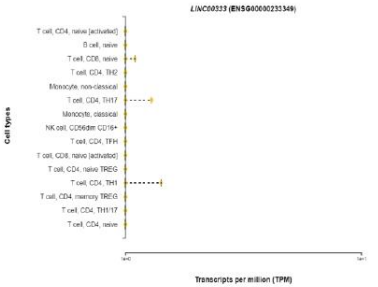


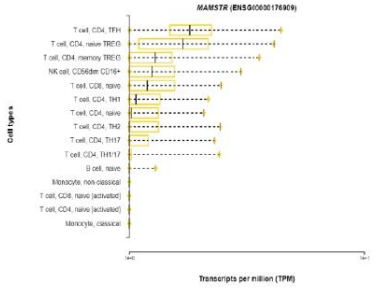


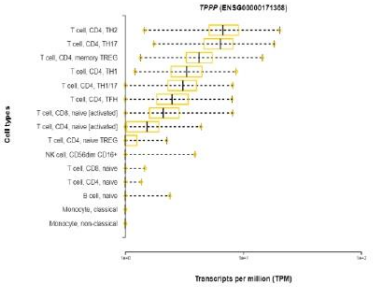


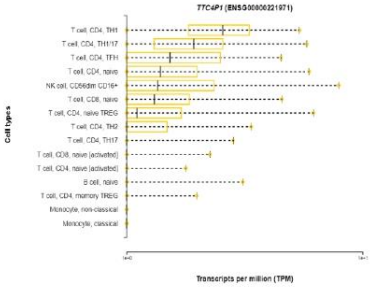


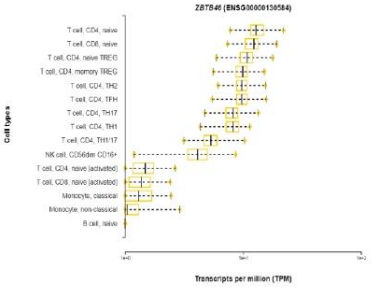












III