

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,*}

^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 (*DDRI*, *GTF2H4*, *VARS2*, *SFTA2* and *DPCRI*) 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

0889-1591/© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

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 hyper-polymorphic 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).

BIGDAGW package v 2.1 in R (Pappas et al., 2016) was used for initial amino acid analysis. Substitution 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-

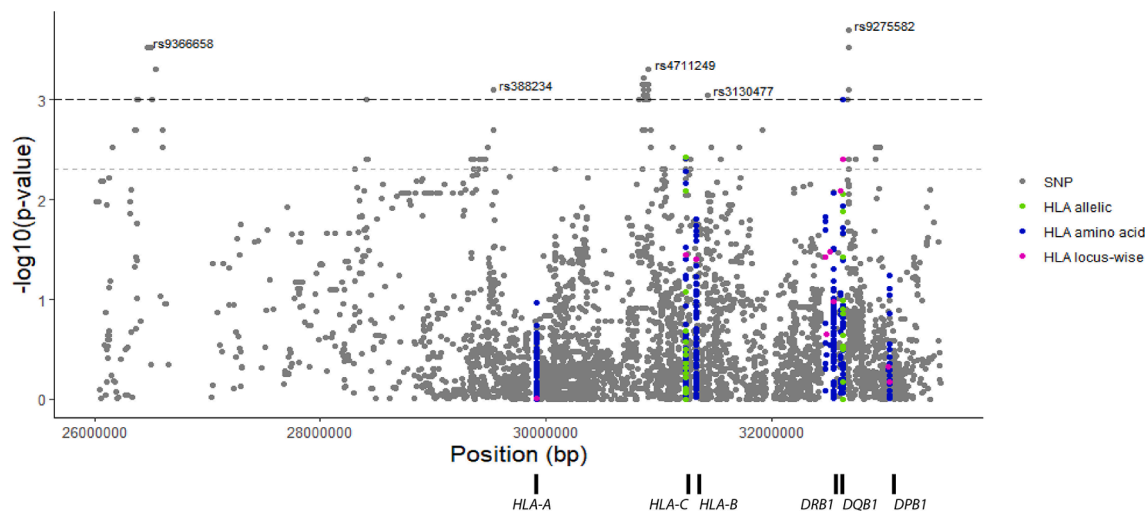


Fig. 1. Single marker association results across the major histocompatibility complex (MHC) in 427 Norwegian Myalgic encephalomyelitis/Chronic fatigue syndrome (ME/CFS) patients and 480 healthy controls. The statistical significance of the association analysis as $-\log_{10}$ of the P-value (y-axis), 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.)

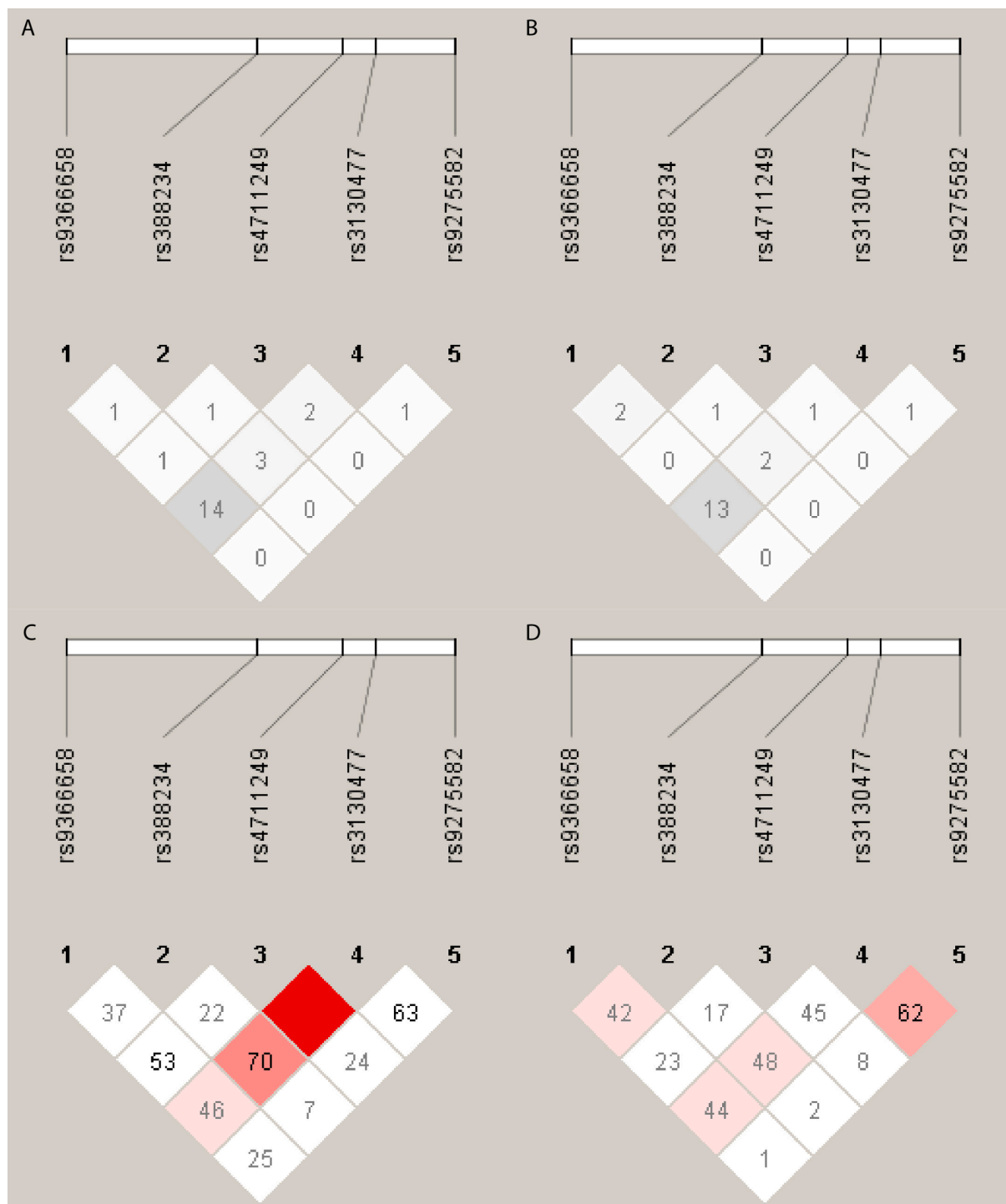


Fig. 2. Linkage disequilibrium (LD) plot for the five tag SNPs rs9366658, rs388234, rs4711249, rs3130477 and rs9275582 measured by r^2 in cases (A) and controls (B), and by D' in cases (C) and controls (D) for 427 Norwegian ME/CFS patients and 480 healthy controls. Values shown in the figures are $100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

SNP	BP	P	P values after conditioning on following SNP				
			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.

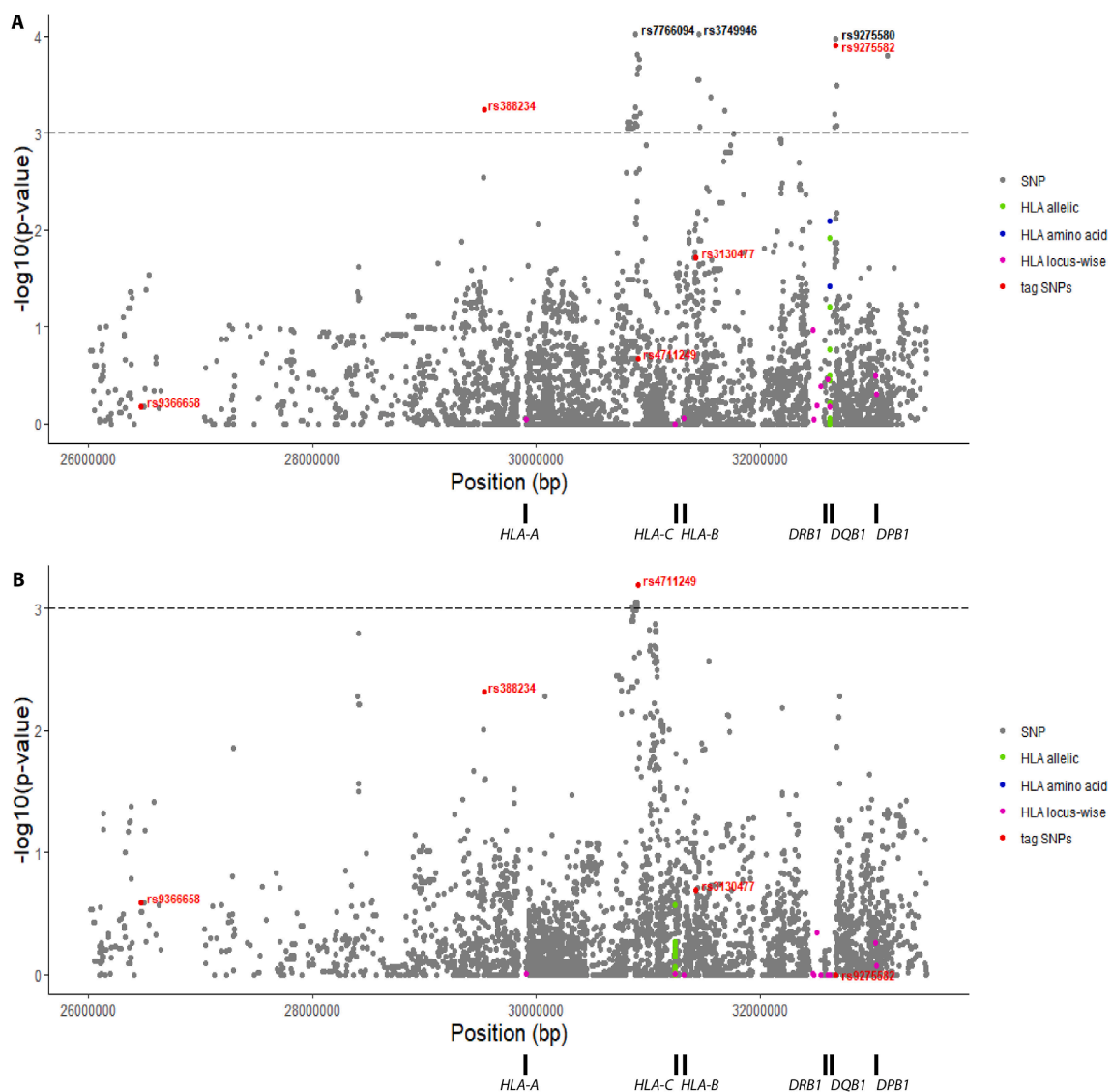


Fig. 3. Single marker association results within the major histocompatibility complex (MHC) in Myalgic encephalomyelitis/Chronic fatigue syndrome (ME/CFS) patients and healthy controls for analyses conditioning on A) *HLA-C* and B) *HLA-DQB1*. The statistical significance of the association analysis as $-\log_{10}$ 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).

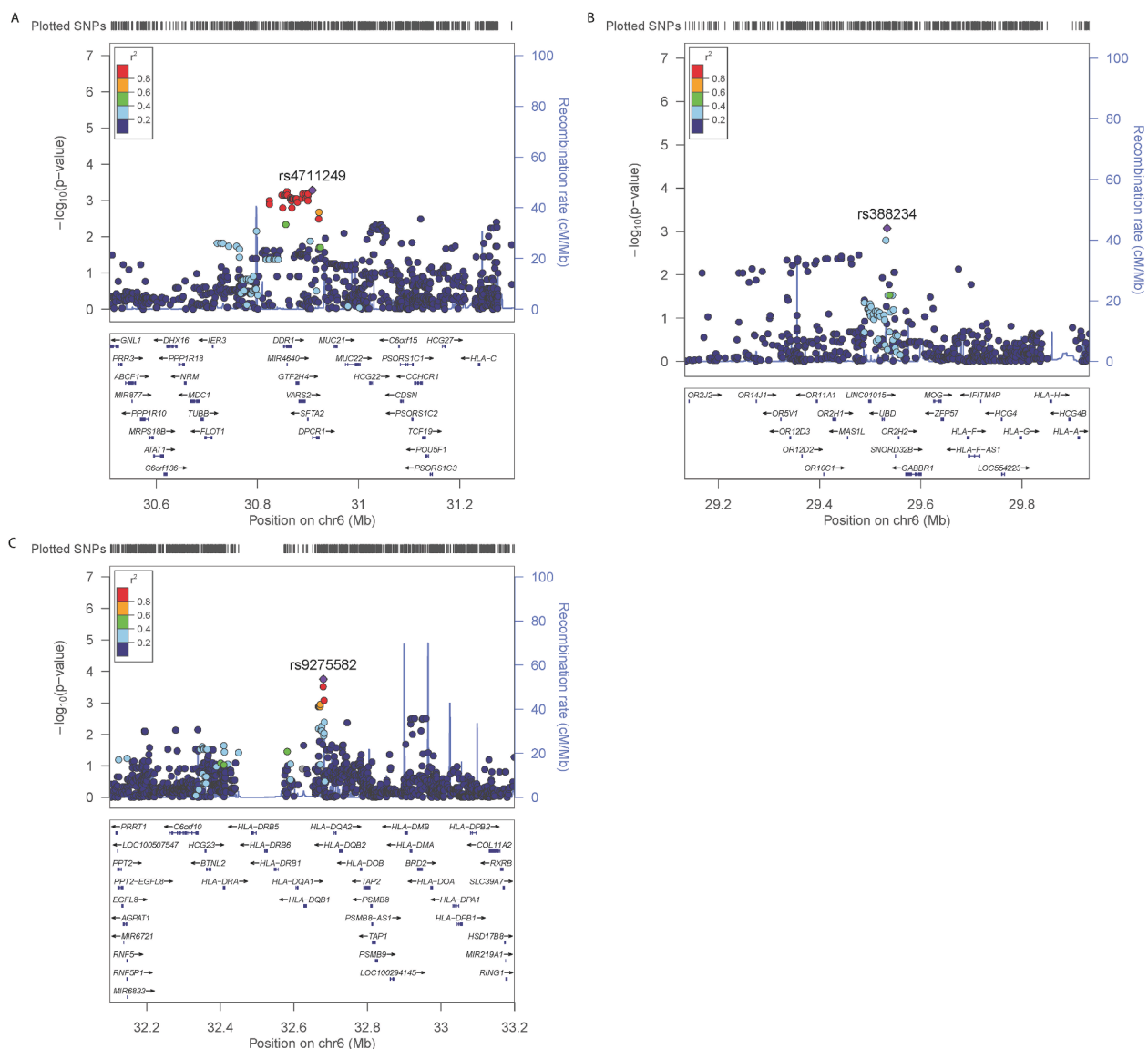


Fig. 4. The significance of the association signals from all SNPs, including their LD with the top SNP, across the disease susceptibility regions tagged by A) rs4711249, B) rs388234 and C) HLA-DQB1. The significance of the association signals, $-\log_{10}P$ -value (y-axis), are plotted against the positions on chromosome 6 shown in base pairs (bp, x-axis). The colour of the dots is indicating r^2 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:			
<i>DQB1*03:01</i>	15.0	13.3	0.3
<i>DQB1*03:03</i>	6.4	4.7	0.1
<i>DQB1*03:09</i>	0	0.1	1
<i>DQB1*03:19</i>	0.1	0	0.5
<i>DQB1*04:02</i>	4.6	4.2	0.7
<i>DQB1*05:03</i>	2.8	1.8	0.1
<i>DQB1*06:01</i>	0	0.5	0.06
<i>DQB1*06:02</i>	17.0	14.9	0.2
<i>DQB1*06:03</i>	7.7	6.5	0.3
<i>DQB1*06:16</i>	0	0.1	1
Residue 57 alanine (57A)	28.8	35.9	0.001
Alleles encoding 57A:			
<i>DQB1*02:01</i>	11.4	13.8	0.1
<i>DQB1*02:02</i>	5.3	5.7	0.7
<i>DQB1*03:02</i>	12.1	16.4	0.009
<i>DQB1*03:04</i>	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 (*DDRI*, *GTF2H4*, *VAR2*, *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 *VAR2* 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 initial 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.decode-me.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 commonalities 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.L., 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 Gabb1 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-Krukke, 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.