



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

Riad Hajdarevic^{a,b,*}, Asgeir Lande^a, Jesper Mehlsen^c, Anne Rydland^{a,b}, Daisy D. Sosa^{d,e}, Elin B. Strand^{f,g}, Olav Mella^h, Flemming Pociotⁱ, Øystein Fluge^h, Benedicte A. Lie^{a,b,j}, Marte K. Viken^{a,j,*}

^a Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway

^b Institute of Clinical Medicine, University of Oslo, Norway

^c Section for Surgical Pathophysiology, Rigshospitalet, University of Copenhagen, Denmark

^d National Advisory Unit on CFS/ME, Norway

^e CFS/ME Center, Oslo University Hospital, Norway

^f Department of Digital Health Research, Division of Medicine, Oslo University Hospital, Oslo, Norway

^g Faculty of Health, VID-Specialized University, Diakonhjemmet Hospital, Oslo, Norway

^h Department of Oncology and Medical Physics, Haukeland University Hospital, Norway

ⁱ Department of Clinical Research, Steno Diabetes Center Copenhagen, Denmark

^j Department of Immunology, Oslo University Hospital and University of Oslo, Oslo, Norway

ARTICLE INFO

Keywords:
ME/CFS
Immunochip
GWAS

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. Heritability is suggested by an increased disease risk in relatives, 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. In the Norwegian discovery cohort, the *TPPP* gene region showed the most significant association (rs115523291, $P = 8.5 \times 10^{-7}$), but we could not replicate the top SNP. However, several other SNPs in the *TPPP* gene identified in the Norwegian discovery cohort showed modest association signals in the self-reported UK biobank CFS cohort, which was also present in the combined analysis of the Norwegian and UK biobank cohorts, *TPPP* (rs139264145; $P = 0.00004$). Interestingly, *TPPP* is expressed in brain tissues, hence it will be interesting to see whether this association, with time, will be verified in even larger cohorts. Taken together our study, despite being the largest to date, could not establish any ME/CFS risk loci, but comprises data for future studies to accumulate the power needed to reach genome-wide significance.

1. 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 (Lim et al., 2020).

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 (Sepúlveda et al., 2019; Sotzny et al., 2018; Rasa et al., 2018; Feiring et al., 2017). We have previously reported association with certain HLA class I and II variants (Lande et al., 2020; Hajdarevic et al., 2021), which also is a hallmark of diseases where the immune system is involved. Furthermore,

* Corresponding authors at: Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway.

E-mail addresses: riad.hajdarevic@medisin.uio.no (R. Hajdarevic), m.k.viken@ous-research.no (M.K. Viken).

<https://doi.org/10.1016/j.bbi.2022.03.010>

Received 26 September 2021; Received in revised form 3 March 2022; Accepted 16 March 2022

Available online 19 March 2022

0889-1591/© 2022 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/>).

comorbidity of ME/CFS with various autoimmune diseases (AIDs) has been observed in a substantial number of patients (Sotzny et al., 2018). In addition, the presence of hereditary components in ME/CFS is supported by excess relatedness (Albright et al., 2011; Lawrie and Pelosi, 1995) (Buchwald et al., 2001) (Carruthers et al., 2003) 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 (Steiner et al., 2020). A candidate SNP study from 2006, of 43 patients, reported four SNPs in the *NR3C1* gene to reach significance ($p < 0.05$; Goertzel et al., 2006). 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 (Schlauch et al., 2016; Smith et al., 2011). Analysis of 353 ME/CFS patients who underwent genotyping by the company “23andme” also highlighted genes in immune pathways (Perez et al., 2019). 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 (Dibble et al., 2020).

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 (Parkes et al., 2013; Matzaraki et al., 2017; Dendrou, 2018; Wang et al., 2015). Furthermore, GWAS have demonstrated that many AIDs share a substantial amount of genetic risk factors with each other (Kochi, 2016; Ricaño-Ponce and Wijmenga, 2013), which lead to the development of the immunochip (Cortes and Brown, 2010). The establishment of genetic risk factors (i.e. $P < 5 \times 10^{-8}$) has been achieved by first conducting independent studies before enabling large consortium studies and meta-analyses of tens of thousands of patients and controls (Visscher et al., 2017). 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 >3000 ME/CFS patients.

2. Material and methods

2.1. Study population

Our discovery cohort comprised 427 Norwegian ME/CFS patients collected from four separate inclusion groups; 1) the Rituximab study (Fluge et al., 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. 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; Carruthers et al., 2003), these four fulfilled the 2010 International Consensus Criteria (Carruthers et al., 2011). For the Norwegian patients, 14.3% (58/407) reported comorbidity with AIDs, with Hashimoto's thyroiditis/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, however, with less and/or different controls and using actual HLA genotypes (Lande et al., 2020; Hajdarevic et al., 2021). Hence, the HLA complex was not investigated in the current study.

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.

2.2. 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 (Beecham, 2013; Liu et al., 2013). 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 12 K 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 (Barrett et al., 2009).

2.3. 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; Das et al., 2016). 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 Fig. 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 ($P < 5 \times 10^{-8}$) 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.

2.4. Databases used for gene expression

The web tool Fuma was used to obtain gene expression data for specific genes in various tissues (Watanabe et al., 2019). In addition, we used Locus Focus (<https://locusfocus.research.sickkids.ca/>) for simple sum calculations, a frequentist colocalization method utilizing the GTEx database v8 gene expression data (Panjwani et al., 2020).

3. 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 Fig. 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 ($P < 1 \times 10^{-5}$; Fig. 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, $P = 8.5 \times 10^{-7}$). The remaining regions displaying $P < 1 \times 10^{-5}$ spanned *UBE2D1* (rs117354281, $P = 8.4 \times 10^{-6}$), *STAB2* (rs11111735, $P = 1.6 \times 10^{-6}$) and *LINC00333* (rs9546628, $P = 3.6 \times 10^{-6}$). 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 Tables 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.

3.1. 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 < 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 < 0.05 , namely rs2582085, rs115523291, rs8108136, rs6089982 (Supplementary Table 4). Hence, none of the tag SNPs showed convincing 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/FUT1* (rs8108136), (Fig. 2), while the odds ratio deviated in the Danish ME/CFS cohort for *TPPP* (rs115523291). Notably, these findings might be due to chance as the confidence intervals crossed $OR = 1$ for one or both replication cohorts. For the remaining SNPs, the results showed much larger differences between the datasets (Supplementary Fig. 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 even the 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).

3.2. 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 Fig. 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 Fig. 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 (Fig. 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/FUT1* (rs28745910; $P = 0.0002$) and *ZBTB46* (rs2777943; $P = 0.0002$). Hence, these regional replication analyses

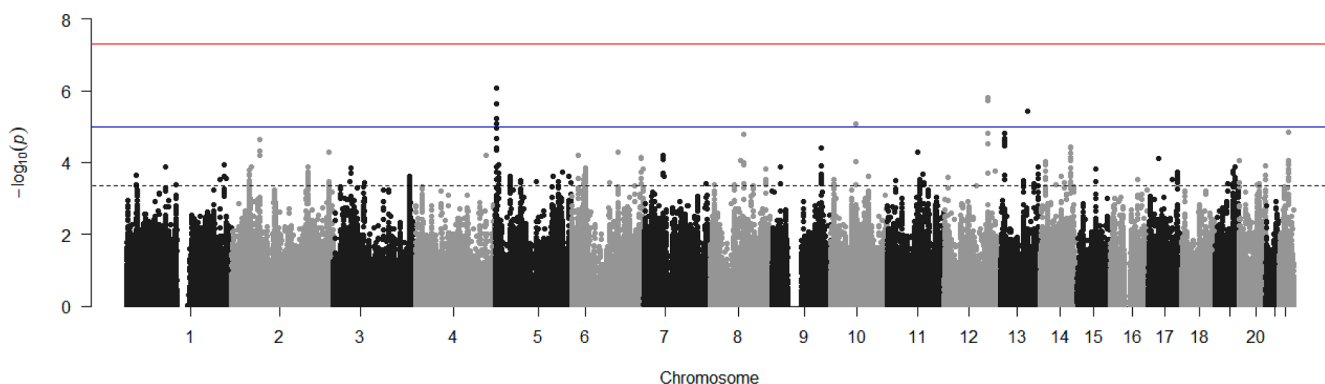


Fig. 1. SNP association results across the 22 autosomes in 427 Norwegian myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) patients and 972 healthy controls. The statistical significance of the association analysis as $-\log_{10}(p)$ (y-axis), is plotted against the chromosomal position of each chromosome in base pairs (bp, x-axis). The red, horizontal line represents a genome wide significance threshold of $P = 5 \times 10^{-8}$, while the blue line represents the suggestive significance level of $P = 1 \times 10^{-5}$ 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). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

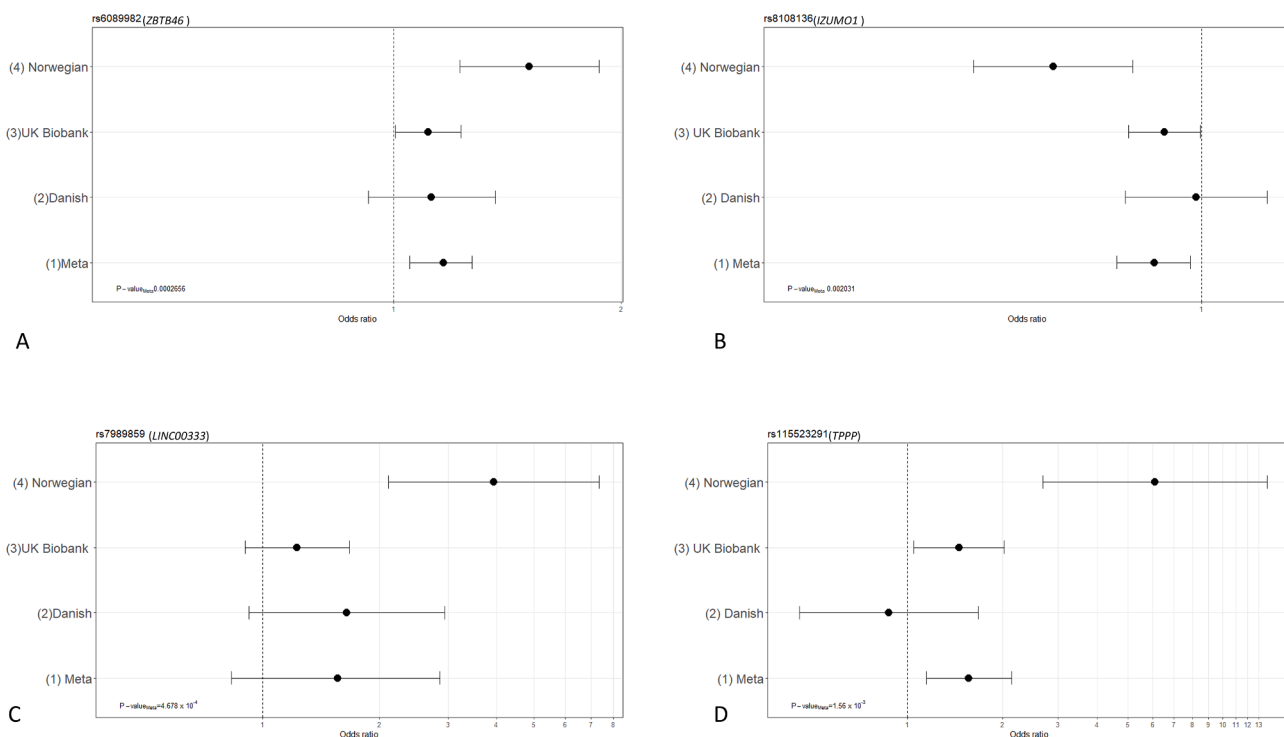


Fig. 2. Odds ratios plots for the tag SNPs rs6089982, rs8108136, rs7989859 and rs115523291 with the odds ratio plotted separately for the Norwegian cohort (N = 427 cases and N = 972 controls), UK biobank cohort (N = 2105 cases and N = 4786 controls) and the Danish cohort (N = 460 cases and N = 1956 controls)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

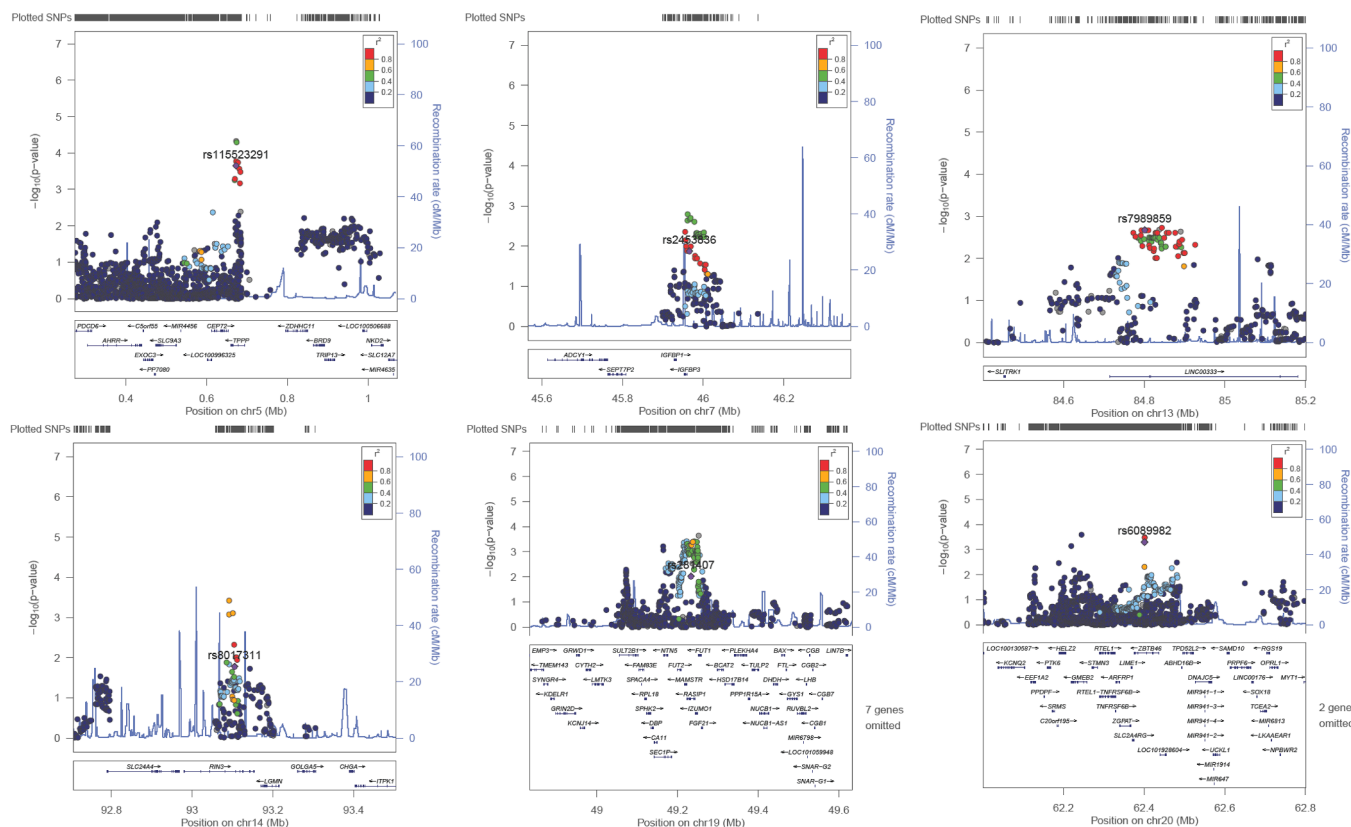


Fig. 3. Locus zoom plots for the regional analyses of the putative ME/CFS associated regions identified in the Norwegian discovery cohort (TPPP, LINC00333, RIN3, IGFBP1/IGFBP3, IZUMO1/FUT1 and ZBTB46). The plots show the meta-analysis results of the Norwegian cohort (N = 427 cases and N = 972 controls) and UK biobank cohort (N = 2105 cases and N = 4786 controls) including imputed SNPs.

pointed out two potentially novel loci, in addition to those revealed in the single tag SNP analysis, namely *RIN3* and *IGFBP1/IGFBP3*.

3.3. 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 ($P < 1 \times 10^{-5}$) and one SNP upstream of *NOX3* (NADPH oxidase 3) on 6q15 reached GWAS level significance (rs77381650, $p = 4.4 \times 10^{-8}$). 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 ($P < 1 \times 10^{-5}$), 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* ($P = 0.02$).

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

Several SNPs encompassing the *TPPP* gene showed association both in the Norwegian (Fig. 4a) and the UK (Fig. 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; $P = 4.7 \times 10^{-5}$), a SNP in strong LD with rs115523291 (Fig. 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 ($P = 8.46 \times 10^{-7}$) 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 (Fig. 4d). Haplotype analyses of rs451979 and rs139264145 showed global associations in both the Norwegian ($P = 9.4 \times 10^{-6}$) and the UK dataset ($P = 0.0005$), but with different individual haplotypes driving the association.

3.5. 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 (Steiner et al., 2020). We found no associations in the Norwegian dataset ($P = 0.9$

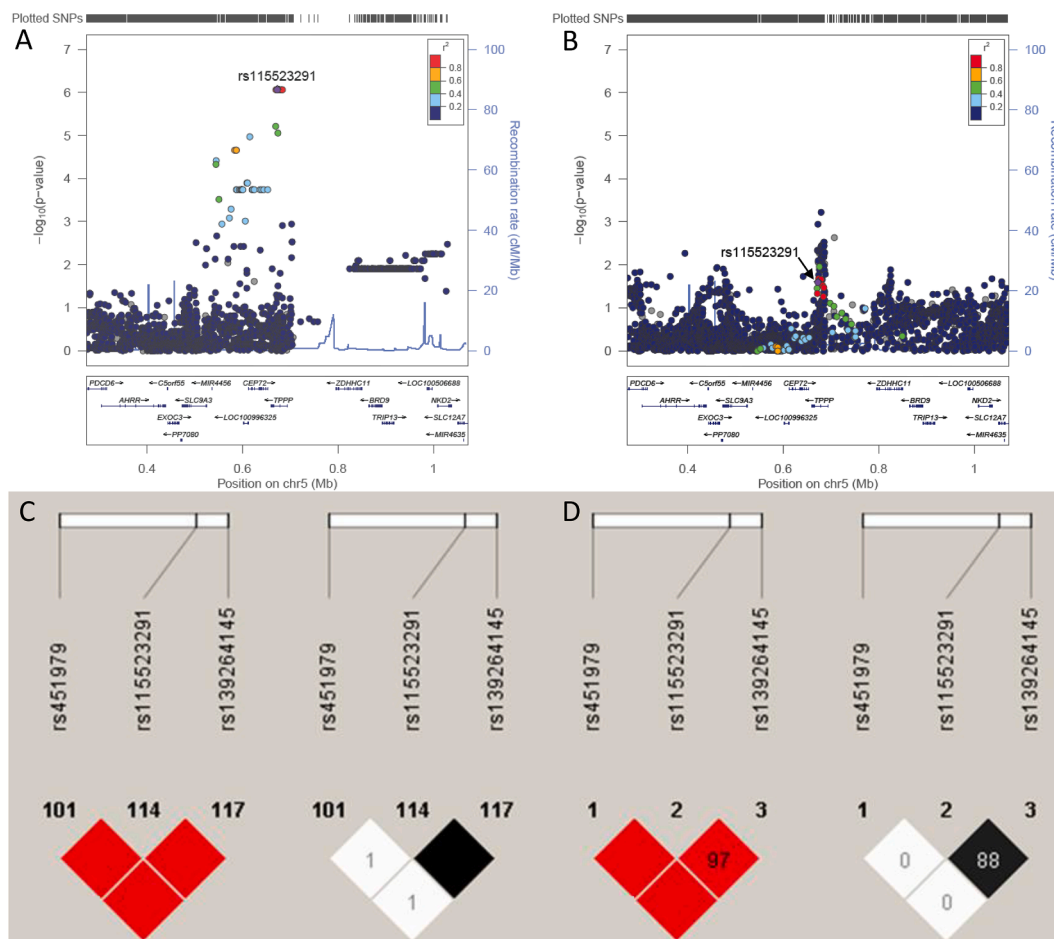


Fig. 4. Locus zoom plots of the tag SNP rs115523291 in (A) the Norwegian cohort (N = 427 cases and N = 972 controls) and (B) in the UK Biobank cohort (N = 2105 cases and N = 4786 controls). Linkage disequilibrium (LD) plot for SNPs rs451979, rs139264145 and rs115523291 measured by r^2 in the Norwegian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

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* ($P \geq 0.004$) and *CTLA4* ($P \geq 0.0007$) genetic regions.

3.6. Gene expression and colocalization

Several of the annotated genes are expressed in brain tissues ([Fig. 5](#)), with *TPPP* and *ZBTB46* showing high expression levels. Few genes appeared to be expressed in whole blood and only *RIN3* showed pronounced expression ([Fig. 5](#)).

However, a simple sum analysis on our top regions for the Norwegian, UK and combined datasets did not reveal any GWAS significant colocalization signals for the regions and relevant tissues (blood, brain, skeletal muscle and nerve) consistent across the cohorts ([Supplementary Table 10](#)).

4. 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. In our second replication strategy, 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. Interestingly, 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*). Nevertheless, it must be emphasized that given our lack of statistical power these observations could be false positive.

In an attempt to overcome the lack of statistical power, we also used the UK Biobank cohort as a discovery cohort and the Norwegian as

replication. However, this approach did not reveal any overlapping significant associations between the two datasets with only ten SNPs of the 228 most associated UK Biobank SNPs also being 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 association signals 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 ([Fu et al., 2019](#)) and has been shown to correlate with shortened disease duration in multiple sclerosis ([Höftberger et al., 2010](#)). 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. ([Frykman et al., 2012](#)). Furthermore, *RIN3* upregulation has recently been reported in Alzheimer mouse models ([Shen et al., 2020](#)). In addition, *IGFBP1* has been associated with Crohn’s disease ([Ye, 2020](#)). Nevertheless, none of our ME/CFS associations reached genome-wide significance and hence needs to be replicated before a role in ME/CFS development can be established. Furthermore, a simple sum colocalization analysis did not reveal any convincing overlap with gene expression differences in relevant tissues, thereby leaving open also the question of which genes represent putative functional risk loci.

We did not replicate previously reported associations with the immunologically important genes, *PTPN22* and *CTLA4* ([Steiner et al., 2020](#)). 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. ([Steiner et al., 2020](#)) 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

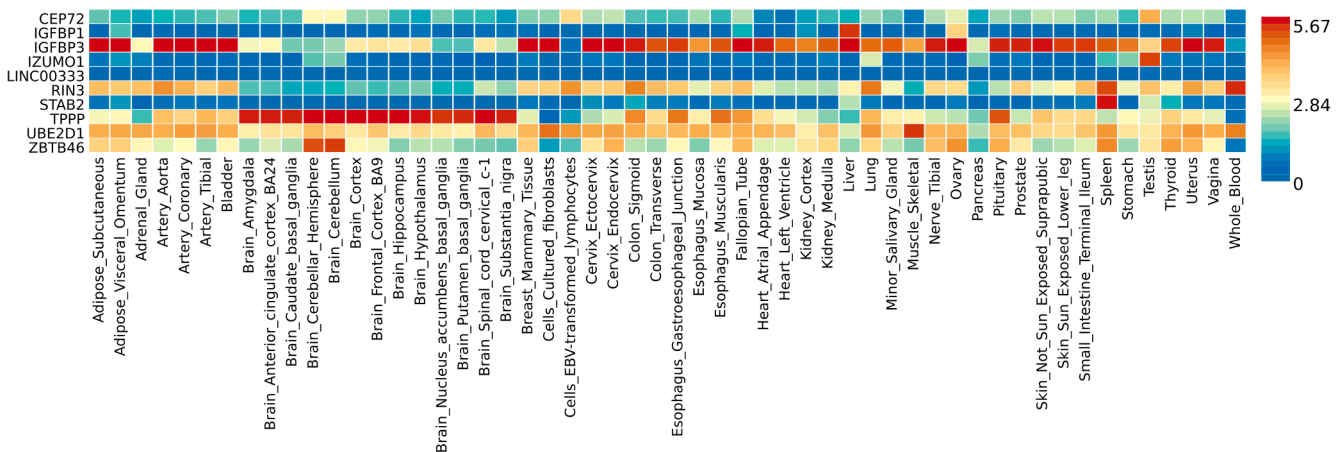


Fig. 5. Gene expression levels in various tissues of candidate genes selected by their proximity to tag and top SNPs.

likely less reliable, however, ideally both studies should have had serological confirmation. We have previously reported an association with the HLA genes in the Norwegian cohort, using actual HLA genotyping (Lande et al., 2020; Hajdarevic et al., 2021). We did not undertake studying the HLA genes in the UK biobank cohort with imputed HLA data in the current study. 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 (Altshuler et al., 2008). 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 (Schlauch et al., 2016; Smith et al., 2011).

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 establish risk loci (Kochi, 2016; Visscher et al., 2017). 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 (Shen and Carlborg, 2013). 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 (<https://www.decode.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.

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. Thank you to Wenche Kristiansen and Hilde Haukeland in 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. This research has been conducted using the UK Biobank Resource under Application Number 43949. 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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bbi.2022.03.010>.

References

- Lim, E.-J., Ahn, Y.-C., Jang, E.-S., Lee, S.-W., Lee, S.-H., Son, C.-G., 2020. Systematic review and meta-analysis of the prevalence of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME). *J. Transl. Med.* 18 (1) <https://doi.org/10.1186/s12967-020-02269-0>.
- 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.2019.02684>.
- 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.
- Rasa, S., Nora-Krukke, Z., Henning, N., Eliassen, E., Shikova, E., Harter, 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>.
- 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.
- Lande, A., Fluge, Ø., Strand, E.B., Flåm, S.T., Sosa, D.D., Mella, O., Egeland, T., Saugstad, O.D., Lie, B.A., Viken, M.K., 2020. Human Leukocyte Antigen alleles associated with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *Sci. Rep.* 10 (1) <https://doi.org/10.1038/s41598-020-62157-x>.
- Hajdarevic, R., Lande, A., Rekeland, I., Rydland, A., Strand, E.B., Sosa, D.D., Creary, L.E., Mella, O., Egeland, T., Saugstad, O.D., Fluge, Ø., Lie, B.A., Viken, M.K., 2021. 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 Behav. Immun.* 98, 101–109.
- 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>.
- Lawrie, S.M., Pelosi, A.J., 1995. Chronic fatigue syndrome in the community prevalence and associations. *Br. J. Psychiatry* 166 (6), 793–797.
- Buchwald, D., Herrell, R., Ashton, S., Belcourt, M., Schmaling, K., Sullivan, P., Neale, M., Goldberg, J., 2001. A twin study of chronic fatigue. *Psychosom. Med.* 63 (6), 936–943.
- 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.
- Steiner, S., Becker, S.C., Hartwig, J., Sotzny, F., Lorenz, S., Bauer, S., Löbel, M., Stittrich, A.B., Grabowski, P., Scheibenbogen, C., 2020. Autoimmunity-related risk variants in PTPN22 and CTLA4 are associated with ME/CFS with infectious onset. *Front. Immunol.* 11 <https://doi.org/10.3389/fimmu.2020.00578>.
- Goertzel, B.N., Pennachin, C., de Souza Coelho, L., Gurbaxani, B., Maloney, E.M., Jones, J.F., 2006. Combinations of single nucleotide polymorphisms in neuroendocrine effector and receptor genes predict chronic fatigue syndrome. *Pharmacogenomics* 7 (3), 475–483.
- Schlauch, K.A., Khaiboullina, S.F., De Meirleir, K.L., Rawat, S., Peterleit, J., Rizvanov, A.A., Blatt, N., Mijatovic, T., Kulick, D., Palotás, A., Lombardi, V.C., 2016. Genome-wide association analysis identifies genetic variations in subjects with myalgic encephalomyelitis/chronic fatigue syndrome. *Transl. Psychiatry* 6 (2), e730.
- Smith, A.K., Fang, H., Whistler, T., Unger, E.R., Rajeevan, M.S., 2011. Convergent genomic studies identify association of GRIK2 and NPAS2 with chronic fatigue syndrome. *Neuropsychobiology* 64 (4), 183–194.
- Perez, M., Jaundoo, R., Hilton, K., Del Alamo, A., Gemayel, K., Klimas, N.G., Craddock, T. J.A., Nathanson, L., 2019. Genetic predisposition for immune system, hormone, and metabolic dysfunction in myalgic encephalomyelitis/chronic fatigue syndrome: a pilot study. *Front. Pediatrics* 7. <https://doi.org/10.3389/fped.2019.00206>.
- Dibble, J.J., McGrath, S.J., Ponting, C.P., 2020. Genetic risk factors of ME/CFS: a critical review. *Hum. Mol. Genet.* 29 (R1), R117–R124.
- Parkes, M., Cortes, A., van Heel, D.A., Brown, M.A., 2013. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat. Rev. Genet.* 14 (9), 661–673.
- Matzaraki, V., Kumar, V., Wijmenga, C., Zhernakova, A., 2017. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biol.* 18 (1) <https://doi.org/10.1186/s13059-017-1207-1>.
- Dendrou, C.A., et al., 2018. HLA variation and disease. *Nat. Rev. Immunol.* 18, 325.
- Wang, L., Wang, F.-S., Gershwin, M.E., 2015. Human autoimmune diseases: a comprehensive update. *J. Intern. Med.* 278 (4), 369–395.

- Kochi, Y., 2016. Genetics of autoimmune diseases: perspectives from genome-wide association studies. *Int. Immunol.* 28 (4), 155–161.
- Ricaño-Ponce, I., Wijmenga, C., 2013. Mapping of immune-mediated disease genes. *Annu. Rev. Genomics Hum. Genet.* 14 (1), 325–353.
- Cortes, A., Brown, M.A., 2010. Promise and pitfalls of the Immuchip. *Arthritis Res. Ther.* 13 (1), 101. <https://doi.org/10.1186/ar3204>.
- Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., Yang, J., 2017. 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* 101 (1), 5–22.
- Fluge, Ø., Bruland, O., Risa, K., Storstein, A., Kristoffersen, E.K., Sapkota, D., Næss, H., Dahl, O., Nyland, H., Mella, O., Reindl, M., 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. <https://doi.org/10.1371/journal.pone.0026358>.
- Fluge, Ø., Rekeland, I.G., Lien, K., Thürmer, H., Borchgrevink, P.C., Schäfer, C., Sørland, K., Abmus, 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>.
- 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 <https://doi.org/10.3389/fmed.2020.00162>.
- Carruthers, B.M., van de Sande, M.I., De Meirleir, K.L., Klimas, N.G., Broderick, G., Mitchell, T., Staines, D., Powles, A.C.P., Speight, N., Vallings, R., Bateman, L., Baumgarten-Austrheim, B., Bell, D.S., Carlo-Stella, N., Chia, J., Darragh, A., Jo, D., Lewis, D., Light, A.R., Marshall-Gradisbik, S., Mena, I., Mikovits, J.A., Miwa, K., Murovska, M., Pall, M.L., Stevens, S., 2011. Myalgic encephalomyelitis: international consensus criteria. *J. Intern. Med.* 270 (4), 327–338.
- 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.
- Liu, J.Z., Hov, J.R., Følseraas, T., Ellinghaus, E., Rushbrook, S.M., Doncheva, N.T., Andreassen, O.A., Weersma, R.K., Weismüller, T.J., Eksteen, B., Invernizzi, P., Hirschfeld, G.M., Gotthardt, D.N., Pares, A., Ellinghaus, D., Shah, T., Juran, B.D., Milkiewicz, P., Rust, C., Schramm, C., Müller, T., Srivastava, B., Dalekos, G., Nöthen, M.M., Herms, S., Winkelmann, J., Mitrovic, M., Braun, F., Ponsioen, C.Y., Croucher, P.J.P., Sterneck, M., Teufel, A., Mason, A.L., Saarela, J., Leppa, V., Dorfman, R., Alvaro, D., Floreani, A., Onengut-Gumuscu, S., Rich, S.S., Thompson, W.K., Schork, A.J., Næss, S., Thomsen, I., Mayr, G., König, I.R., Hveem, K., Cleyneen, I., Gutierrez-Achury, J., Ricaño-Ponce, I., van Heel, D., Björnsson, E., Sandford, R.N., Durie, P.R., Melum, E., Vatn, M.H., Silverberg, M.S., Duerr, R.H., Padyukov, L., Brand, S., Sans, M., Annese, V., Achkar, J.-P., Boberg, K. M., Marschall, H.-U., Chazouillères, O., Bowliu, C.L., Wijmenga, C., Schrupf, E., Vermeire, S., Albrecht, M., Rioux, J.D., Alexander, G., Bergquist, A., Cho, J., Schreiber, S., Manns, M.P., Färkkilä, M., Dale, A.M., Chapman, R.W., Lazaridis, K.N., Franke, A., Anderson, C.A., Karlsen, T.H., 2013. Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat. Genet.* 45 (6), 670–675.
- Barrett, J.C., Clayton, D.G., Concannon, P., Akolkar, B., Cooper, J.D., Erlich, H.A., Julier, C., Morahan, G., Nerup, J., Nierras, C., Plagnol, V., Pociot, F., Schuilenburg, H., Smyth, D.J., Stevens, H., Todd, J.A., Walker, N.M., Rich, S.S., 2009. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.* 41 (6), 703–707.
- 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.
- Watanabe, K., Umičević Mirkov, M., de Leeuw, C.A., van den Heuvel, M.P., Posthuma, D., 2019. Genetic mapping of cell type specificity for complex traits. *Nat. Commun.* 10 (1) <https://doi.org/10.1038/s41467-019-11181-1>.
- Panjwani, N., Wang, F., Mastromatteo, S., Bao, A., Wang, C., He, G., Gong, J., Rommens, J.M., Sun, L., Strug, L.J., Perthea, M., 2020. LocusFocus: Web-based colocalization for the annotation and functional follow-up of GWAS. *PLoS Comput. Biol.* 16 (10), e1008336. <https://doi.org/10.1371/journal.pcbi.1008336>.
- Fu, M.-M., McAlear, T.S., Nguyen, H., Oses-Prieto, J.A., Valenzuela, A., Shi, R.D., Perrino, J.J., Huang, T.-T., Burlingame, A.L., Bechstedt, S., Barres, B.A., 2019. The Golgi outpost protein TPPP nucleates microtubules and is critical for myelination. *Cell* 179 (1), 132–146.e14.
- Höftberger, R., Fink, S., Aboul-Enein, F., Botond, G., Olah, J., Berki, T., Ovadi, J., Lassmann, H., Budka, H., Kovacs, G.G., 2010. Tubulin polymerization promoting protein (TPPP/p25) as a marker for oligodendroglial changes in multiple sclerosis. *Glia* 58 (15), 1847–1857.
- Frykman, S., Teranishi, Y., Hur, J.-Y., Sandebring, A., Goto Yamamoto, N., Ancarcrona, M., Nishimura, T., Winblad, B., Bogdanovic, N., Schedin-Weiss, S., Kihara, T., Tjernberg, L.O., 2012. Identification of two novel synaptic γ -secretase associated proteins that affect amyloid β -peptide levels without altering Notch processing. *Neurochem. Int.* 61 (1), 108–118.
- Shen, R., Zhao, X., He, L., Ding, Y., Xu, W., Lin, S., Fang, S., Yang, W., Sung, K., Spencer, B., Rissman, R.A., Lei, M., Ding, J., Wu, C., 2020. Upregulation of RIN3 induces endosomal dysfunction in Alzheimer's disease. *Transl. Neurodegener.* 9 (1) <https://doi.org/10.1186/s40035-020-00206-1>.
- Ye, B.D., et al., 2020. Association of FUT2 and ABO with Crohn's disease in Koreans. *J. Gastroenterol. Hepatol.* 35 (1), 104–109.
- Altshuler, D., Daly, M.J., Lander, E.S., 2008. Genetic mapping in human disease. *Science* 322 (5903), 881–888.
- Shen, X., Carlborg, O., 2013. Beware of risk for increased false positive rates in genome-wide association studies for phenotypic variability. *Front. Genet.* 4, 93.