Genetic Risk Factors for ME/CFS Identified using Combinatorial Analysis

Sayoni Das*, Krystyna Taylor*, James Kozubek, Jason Sardell, Steve Gardner*†

PrecisionLife Ltd, Long Hanborough, Oxford, UK

* These authors contributed equally.
† Corresponding author: steve@precisionlife.com

Abstract

Background: Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is a debilitating chronic disease that lacks known pathogenesis, distinctive diagnostic criteria, and effective treatment options. Understanding the genetic (and other) risk factors associated with the disease would begin to help alleviate some of these issues for patients. **Methods:** We applied both GWAS and the PrecisionLife combinatorial analytics platform to analyze ME/CFS cohorts from UK Biobank, including the Pain Questionnaire cohort, in a case-control design with 1,000 cycles of fully random permutation. The results from this study were supported by a series of replication and cohort comparison experiments, including use of a disjoint Verbal Interview cohort also derived from UK Biobank, and results compared for reproducibility. **Results:** Combinatorial analysis revealed 199 SNPs mapping to 14 genes, that were significantly associated with 91% of the cases in the ME/CFS population. These SNPs were found to stratify by shared cases into 15 clusters (communities) made up of 84 high-order combinations of between 3-5 SNPs. *p*-values for these communities range from 2.3×10^{-10} to 1.6×10^{-72}. Many of the genes identified are linked to the key cellular mechanisms hypothesized to underpin ME/CFS, including vulnerabilities to stress and/or infection, mitochondrial dysfunction, sleep disturbance and autoimmune development. We noted similarities with genes associated with multiple sclerosis and long COVID, which share some symptoms and potentially a viral infection trigger with ME/CFS. **Conclusions:** This study provides the first detailed genetic insights into the pathophysiological mechanisms underpinning ME/CFS and offers new approaches for better diagnosis and treatment of patients.

Keywords: ME/CFS, Combinatorial Analytics, Patient Stratification, Biomarkers, Novel Targets, Precision Repositioning

Background

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is a debilitating chronic disease that presents with diverse symptoms including post-exertional malaise, chronic pain, and cognitive impairment[1]. It affects approximately 0.2% of the UK population[2]. There are currently no approved disease modifying therapies for ME/CFS, and patients are managed via prescription of drugs and other therapies for symptomatic relief, including pain relief, anti-depressants, and cognitive behavioural therapy[3].

The breadth of symptoms and severities experienced by ME/CFS patients is likely indicative of the heterogeneous nature of the disorder, with a variety of metabolic, immunological, neuroendocrine and central nervous system dysfunctions underlying an individual patient's pattern of onset and development of the disease. ME/CFS development has been associated with prior viral infection such as with Epstein-Barr Virus (EBV)[4] and other pathogens[5,6,7,8], however there is also evidence that stress and non-viral infection may also contribute to triggering ME/CFS onset[9].

The multi-factorial spectrum of ME/CFS triggers and symptoms[10] invites the question whether ME/CFS may represent multiple patient subgroups with a range of potentially overlapping underlying biological drivers. If so, better characterization of the etiology of disease in these subgroups may lead to improved understanding of ME/CFS and identification of personalized treatments that are most effective for specific subgroups.

Previous ME/CFS population studies have performed Genome-Wide Association Studies (GWAS) with the aim of identifying significant genetic factors underlying disease risk. While there is a demonstrable heritable component to the disease[11], no significant single-gene association to ME/CFS has been identified using analysis of whole exome sequences, and given the limited statistical power associated with the small ME/CFS genetic datasets available, GWAS approaches have been unable to detect disease-associated SNPs that exhibit sufficiently large effect sizes across the whole of the patient population[12].

ME/CFS is clearly not a simple monogenic disease caused by single nucleotide variants (SNVs) with large effect sizes but is likely caused by complex interactions of many genetic, epidemiological and environmental factors that GWAS-based approaches are not able to fully identify. This requires a different analytical approach.

Combinatorial Analysis

Although GWAS has helped to transform the treatment of many relatively monogenic diseases by revealing clinically relevant single SNP genetic associations, it has been less successful in complex, chronic diseases. These are more polygenic and heterogeneous with patients presenting in a spectrum, and they may include high-resolution signals such as disease-associated variants occurring within linkage disequilibrium (LD) blocks[13,14].

Notably, inclusion of patients with different etiologies under the same “case” classification weakens SNP-disease associations in GWAS, causing the method to potentially overlook the genetic variants responsible for disease in subsets of the population. More fundamentally, GWAS is not designed to detect epistatic and other non-linear effects caused by the interactions of multiple variants. As such it struggles to identify variants that are strongly associated with different patient subgroups in a heterogeneous patient population with multiple diverse disease etiologies that may be further influenced by non-linear interactions across and between multiple genes and transcription/expression control regions.

This however is exactly the challenge presented by ME/CFS and other complex, chronic diseases. Understanding of how the range of disease etiologies affects different patient subgroups requires the identification of combinations of SNPs (and other clinical, transcriptomic and/or epidemiological or environmental features) that together are co-associated with a specific phenotype.

The PrecisionLife combinatorial analysis platform enables hypothesis-free identification of such high-order combinatorial multi-modal features (known as disease signatures) at scale on modest computational hardware. These combinatorial disease signatures capture both linear and non-linear effects of genetic and molecular interaction networks in a way that is complementary to GWAS analysis[15].

The combinatorial approach is more sensitive than GWAS, enabling identification of novel genetic associations and mechanisms that may only be relevant to a subgroup of patients, leading to more validated associations than GWAS when analyzing the same datasets. This approach has been validated in multiple disease studies both by the authors and collaborators, in some cases using *in vitro* and *in vivo* disease assays to demonstrate novel target genes’ disease modification potential, and in others by the presence in pharmaceutical companies’ R&D pipelines of drug programs targeting mechanisms that were identified by combinatorial analysis, but which could not be found using GWAS on available patient datasets[16,17,18].

For example, using combinatorial analysis we were first to report the association of 156 loci and 68 genes with the risk of developing severe COVID-19[19]. This analysis was run on just 725 patients and 1,450 controls from UK Biobank, and contrasts with the 11 and 13 loci discovered using a GWAS approach respectively by 23andMe (16,500 patients/controls)[20] and COVID-19 HGI consortium (over 2,000,000 patient/controls)[21] in similar studies. Of the 68 genes that we reported, 48 have subsequently been associated by other groups with the disease using methods including single-cell analysis and transcriptomic profiling (unpublished literature analysis - June 2022).

Materials and Methods

We analyzed genotype data from 2,382 patients reporting an ME/CFS diagnosis in the UK Biobank Pain Questionnaire[22] matched against 4,764 controls in a case:control study design in the PrecisionLife platform.

Data Sources

ME/CFS patients with a (self-reported) clinical diagnosis in UK Biobank's Pain Questionnaire (Data-field 120010) were identified, of whom over 90% were of European genetic ancestry (Figure 10 in Supplementary Data). Given this proportion, only ME/CFS patients of European genetic ancestry were selected as the case cohort for this study. To ensure properly characterized control subjects, individuals were selected who had no evidence in the Hospital Episodes Statistics (HES), primary care, or self-reported data fields indicating diagnoses of chronic fatigue, post-exertional malaise, post-viral fatigue syndrome or myalgia (see Table 8 in Supplementary Data). To avoid potential confounding, controls meeting these criteria were matched by genetic ancestry and gender against the cases in a 2:1 ratio (and this was repeated with a separate 4:1 ratio study). Because of the narrow age range of UK Biobank participants, strict age matching was not necessary, as cases and controls exhibit very similar age distributions (Figure 12a in Supplemental Data).

Data about individuals' diagnosis with autoimmune disease (including multiple sclerosis and fibromyalgia), self-reported emotional or physical stress, and exposure to potential viral triggers (such as EBV seropositivity) were used to compare ME/CFS cases against the remainder of the UK Biobank to identify any significant differences between the two cohorts that could be associated with ME/CFS onset (see Figure 2).

After quality control (see Genotype Quality Control in Supplementary Data), the Pain Questionnaire dataset was comprised of 2,382 ME/CFS cases, 4,764 controls and 519,337 SNPs on autosomal chromosomes. Approximately 71% of cases (n=1,695) were women (**Error! Reference source not found.** in Supplementary D ata) vs the UK Biobank distribution of 54.4%. The age and body-mass index (BMI) distributions of cases and controls were similar (Figure 12b in Supplementary Data).

Methods

We applied the PrecisionLife platform to the various ME/CFS case-control datasets to identify combinations of SNP genotypes that when observed together in a sample were strongly associated with the development of ME/CFS. The PrecisionLife platform uses a unique data analytics framework that enables efficient combinatorial analysis of large, multi-dimensional participant datasets. Navigating this data space allows for the identification of combinations of features that are significantly associated with groups of cases in a case-control dataset. The PrecisionLife combinatorial analysis is hypothesis free, involving a four-stage mining, validation, evaluation and annotation process.

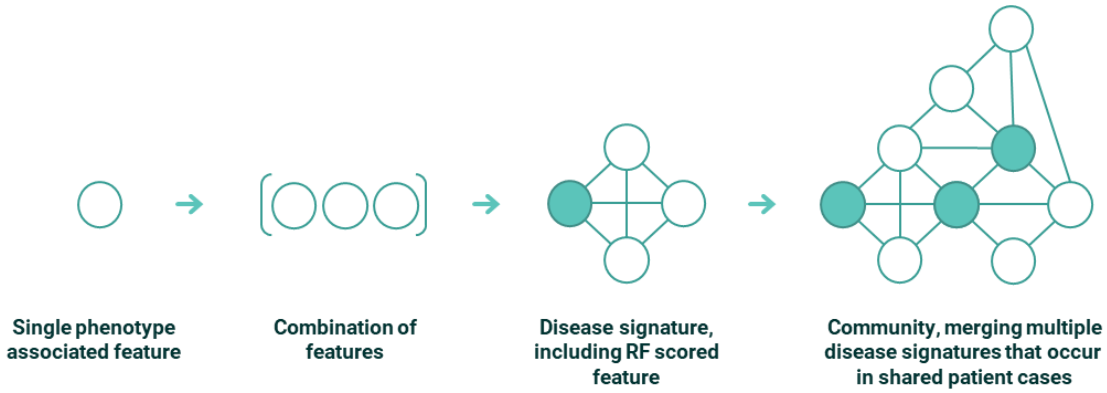


Figure 1. Conceptual representation of features, combinations, disease signatures and communities used to build up the disease architecture in the PrecisionLife combinatorial methodology. In the case of the ME/CFS study all features were SNP genotypes, but other feature types, e.g., a patient's expression level of a specific protein, medication history or their eosinophil level, can also be used.

The PrecisionLife platform identifies combinations of feature states in 'layers' of increasing combinatorial complexity, i.e., singletons, pairs, triplets etc. A feature could for example be a SNP, and a feature state would consist of the SNP's base index and its genotype, which would typically be encoded ordinally as {0, 1, 2} for homozygous major allele, heterozygous minor allele, homozygous minor allele respectively. The platform has considerably more flexibility of representation (including alternate genotype encodings, extended genetic models, polyploidy and quantitative values) if required by the feature or dataset being analyzed.

In the mining phase, combinations of feature states that are overrepresented (using a Z-score or Fisher's Exact test) in cases are identified and validated (Table 1). Multiple feature states are combined iteratively until no additional features can be added that will improve the score. Combinations of feature states that have high odds ratios, low *p*-values ($p < 0.05$) and high prevalence (>5%) in cases are prioritized. The mining process is repeated across up to 1,000 cycles of fully randomized permutation of the case:control labels of all individuals in the dataset, keeping the same parameters and case-control ratio.

In the validation phase, all combinations generated by the original mining run and each of the random permutation iterations of the dataset are compared. These combinations are validated using network properties such as minimum prevalence (number of cases represented, in this case >5%) as the null hypothesis when compared with the combinations generated by the random permutations. Combinations that appear in the random permutations above a specified FDR threshold (Benjamini-Hochberg FDR of 0.05) after multiple testing correction[23] are considered to be random and eliminated. Combinations passing these tests are reported as validated disease signatures.

Table 1: Summary of thresholds used in the mining phase of the PrecisionLife combinatorial analysis on the Pain Questionnaire cohort.

Layer	Minimum Z-Score	Minimum Cases (%)	Minimum Risk Ratio	No. of Reported Disease Signatures
1 (singleton)	1.96	5%	2.00	0
2 (pair)	2.85	5%	2.00	184
3 (triple)	3.83	5%	2.34	45,711
4 (quadruple)	4.21	5%	3.10	49,894
5 (quintuple)	4.33	5%	3.97	44,426

The validated disease signatures are then evaluated. The features (which in this case only consisted of SNPs due to the limited available dataset) shared by multiple disease signatures (known as 'critical' SNPs) are identified. Critical SNPs, which can be thought of as the canonical features of a cluster comprised of overlapping disease signatures, are then scored using a Random Forest (RF) algorithm in a 5-fold cross-validation framework to evaluate the accuracy with which they predict the observed case-control split in a dataset (minimizing Gini impurity or the probability of misclassification).

We use RF scores in similar ways to rank critical SNPs and by association the genes to which they map via the process described in the Functional Genomics Annotation section below. Disease signatures comprising high RF scoring critical SNPs (and their genes) are then mapped to the cases in which they were found, and additional clinical data (such as blood biochemistry data, comorbidity ICD-10 codes and medication history) is used to generate a patient profile for each combinatorial disease signature.

Finally, a merged network (disease architecture) view is generated by clustering all validated disease signatures based on their co-occurrence in patients in the dataset, and annotation of the validated SNPs, genes and the druggability of targets is performed using a semantic knowledge graph (see Functional Genomics Annotation section).

The PrecisionLife platform generated statistically significant ME/CFS associated signatures containing up to five SNPs for each cohort. Each ME/CFS dataset analysis took around 7 days (168) hours to complete, running on a server with 64 CPU cores and 4x Nvidia GPUs.

Replication and Validation

No similarly sized ME/CFS cohort is currently available for use as an independent replication study cohort. We therefore used two alternate approaches to validate the results from the Pain Questionnaire study.

In the first approach, we performed the 1,000 random permutation tests on each combinatorial disease signature by randomly shuffling cases and controls in the Pain Questionnaire dataset and calculating a permutation test score (P1000) for all observed SNP combinations across the full range of combinatorial order. The P1000 score of a combinatorial disease signature indicates the frequency of detection of similarly associated combinatorial features in the 1,000 randomized permutations, as measured by odds ratio and number of cases possessing the feature. Any feature where P1000 is less than 50 (i.e., 5%) is usually considered significant.

In a second approach, we generated a new CFS case population from UK Biobank comprising of individuals with a clinical CFS diagnosis (Data-Field 20002, Coding 1482) reported during Verbal Interview, and compared the results from it with the Pain Questionnaire cohort. The Verbal Interview case population had been analyzed in a recent GWAS study[24]. As the Verbal Interview cohort had 735 individuals in common with the Pain Questionnaire cohort (Figure 17 in Supplementary Data), these overlapping cases were removed to create a disjoint Verbal Interview dataset (cases = 1,273 and controls = 4,137 after QC) of European ancestry with gender (and ancestry) matched controls.

The disjoint Verbal Interview dataset was also analyzed through the PrecisionLife platform to investigate the extent to which the results generated from the original Pain Questionnaire cohort could be replicated in this second cohort. Two issues contribute to limit the degree of overlap that can be expected in a fully independent analysis of the second cohort.

Firstly, as the combinatorial search space is vast, the sampling of that space is likely to be materially incomplete, which will contribute to a potential high rate of false negatives, i.e. true associations that were not tested or reported due to random sampling bias. A separate more systematic sampling of the space will be run in future studies, but this method was not available for this study. Secondly, while the two UK Biobank datasets are based on different clinical diagnoses, the assignment by a GP of either a CFS or ME/CFS diagnosis is highly variable and cannot be relied upon to distinguish the populations in a clinically meaningful manner.

Because of the combination of these two factors, it is likely that the two studies will differ significantly in the reported similarity of their genetic associations and clinical characteristics.

We therefore limited the search space for the analysis of the disjoint Verbal Interview cohort by testing only combinations involving the 199 SNPs identified in the Pain Questionnaire cohort. Limiting the search space to combinations involving these critical SNPs enables us to assess the level of replication of the ME/CFS genetic signal in the second dataset by eliminating the unavoidable sampling bias arising from differences in the heterogeneous patient populations exacerbated by the small numbers of case available in the huge search space.

Functional Genomics Annotation

We mapped all SNPs identified in the disease signatures using an annotation cascade process to the human reference genome (GRCh37)[25] to give the best estimate of the gene(s) likely to be associated with the SNP. Disease-associated SNPs that lie within coding regions of gene(s) were assigned directly to the corresponding gene(s). Remaining SNPs that lie within 2kb upstream or 0.5kb downstream of any gene(s) were mapped to the closest gene(s) within this region. The potentially causality and druggability of these genes were evaluated in later steps.

We investigated additional gene assignments for the identified SNPs using publicly available eQTL[26] and/or chromatin interaction data[27] (see Supplementary Table 11). Genes with at least one cis-eQTL SNP at a false

discovery rate (FDR) of ≤0.05, with expression differences of that gene in single brain tissues or whole blood were reported[26].

Additionally, promoter capture Hi-C (pcHi-C) interactions that were significantly associated in brain tissues and blood cells by Jung et. al [27] were used to generate gene assignments. Due to the uncertainty about the relevant cells and tissues affected in ME/CFS etiology, genes assigned by either eQTL or chromatin interaction data were not specifically prioritized for further analysis (as they might be in other studies) to avoid capturing any spurious associations from non-trait-related tissues[26]. Genes that could be additionally mapped using only eQTL or HiC data from the 25 critical SNPs were however observed and reported in Supplementary Table 11, although these were not further evaluated. The direction of association of any eQTLs associated with the disease phenotype was however noted.

Critical SNPs (see Methods section) were assigned an RF score, describing how well the SNP genotype combinations predict the observed case-control split. We used these scores to rank the critical SNPs to reflect the relative importance of the SNP and its combinations. The genes assigned to the critical SNPs were prioritized on the basis of the cumulative sum of their associated SNP scores to identify the most clinically relevant targets, as the critical SNPs are those observed to have markedly higher association with the disease.

We used a semantic knowledge graph derived from over 50 public and private data sources to annotate the prioritized genes (see Supplementary Table 12). This included information from a variety of data sources including basic genomic context, tissue expression, chemical tractability, biological function and associated scientific literature. We tested each of the genes identified against the 5Rs criteria[28] of early drug discovery, to form and validate hypotheses for their mechanism of action and impact on the disease phenotype.

Patient Stratification

The output disease signatures generated by the PrecisionLife platform contain metadata including the indices of all the cases (and controls) in which they were found. The available phenotypic and clinical data for the relevant cases were used to evaluate patient profiles associated with each of the disease signatures. This was based on the observed enrichment of an attribute or phenotype for a particular group of patients (for example association with a prioritized gene) compared against the entire case population. Statistical significance was calculated using the two proportions Z-test for categorical variables such as gender and co-morbidities, whereas we used the Mann-Whitney U test for continuous variables such as measurements of metabolic biomarkers. *p*-values corrected for multiple-testing using the Benjamini-Hochberg method to control the FDR were also reported.

Results

UK Biobank ME/CFS (Pain Questionnaire) Cohort Characteristics

We identified significant differences in a variety of covariates (listed in Cohort Analysis section in Supplementary Data) between the ME/CFS case population, and the control group, and the remaining individuals in UK Biobank (Figure 2). The figure shows the percentage of individuals in each group who are positive for each covariate. To test for significance, we calculated 95% confidence intervals using bootstrapping (sampling with replacement) for 1,000 iterations.

The greatest difference between the ME/CFS population in this study and the remainder of the UK Biobank was the significantly higher proportion of individuals reporting mental distress and stressful events such as illness, injury, and bereavement. Individuals with ME/CFS in this study were also slightly more likely to present with at least one autoimmune disease, with the greatest co-association with other myalgia and fatigue-associated conditions like multiple sclerosis and fibromyalgia. It is however impossible to rule out a level of misdiagnosis in these complex conditions.

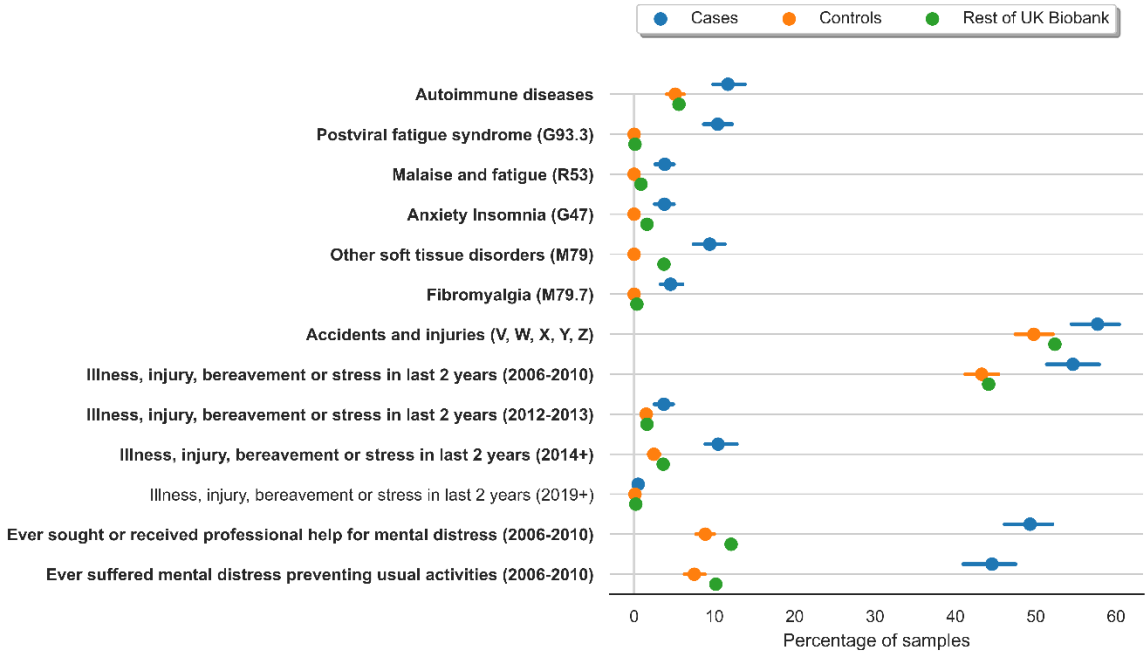


Figure 2. Forest plot showing percentage of individuals in cases, controls and rest of the individuals in UK Biobank who report each covariate along with 95% confidence interval generated using bootstrapping for 1,000 iterations. **Bold** covariate label indicates $p<0.001$, regular label indicates $p<0.01$

Combinatorial Analysis

The ME/CFS Pain Questionnaire cohort (2,382 cases, 4,764 controls) was used to perform a standard GWAS case-control association analysis using PLINK[29]. No SNPs were reported to be significant below a genome-wide significance threshold of $p<5 \times 10^{-8}$ (Figure 3).

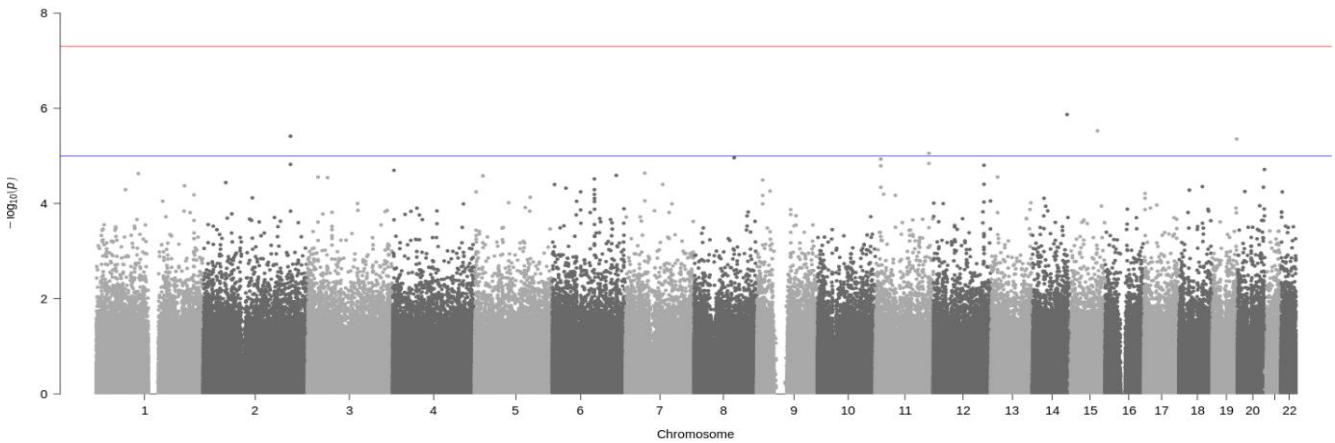


Figure 3: Manhattan plot generated using PLINK of genome-wide *p*-values of association for the Pain Questionnaire cohort (n = 7,146 where cases=2,382 and controls=4,764). The horizontal blue and red dashed lines represent the genome-wide significance values of $p<1 \times 10^{-5}$ and $p<5 \times 10^{-8}$ respectively.

Combinatorial SNP analysis performed using the PrecisionLife platform on the same Pain Questionnaire dataset generated 84 statistically validated combinations of 199 SNPs that together are strongly associated with ME/CFS diagnosis (Table 2, Figure 4). None of the SNPs identified were observed to be in linkage disequilibrium (LD) (Figure 13 in Supplementary Data). 192 SNPs identified in the disease signatures were in non-coding regions of the genome and 9 (7 missense and 2 synonymous) SNPs were identified in the coding regions (Figure 14 in Supplementary Data).

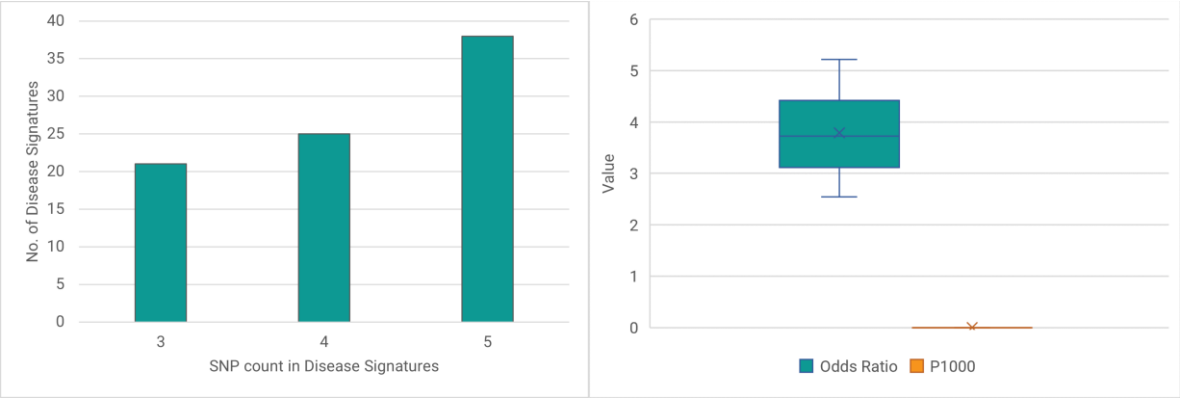
All SNPs were found in combinations with 3 or more SNPs, and so would not have been found using standard GWAS analysis methods (Figure 4a). No single SNPs or SNP pairs were reported as significant by the method.

Table 2: Summary of the results of PrecisionLife combinatorial analysis run on the Pain Questionnaire cohort.

Validated Disease Signatures (SNP Combinations)	84
Total Disease Associated SNPs in Disease Architecture	199
Critical SNPs	25
Prioritized Genes	14
Cases represented by Disease Architecture (%)	91%

As a negative validation check, runs using the same mining and validation parameters as used above were performed on 7,146 random samples, comprising 2,382 UK Biobank randomly sampled participants as 'cases' compared against 4,764 randomly sampled 'controls'. This analysis yielded no significant results.

These 84 combinatorial disease signatures all had P1000 values of 0, indicating they were not detected in any of the 1,000 random permutation runs, and are therefore very unlikely to result from random chance (Figure 4b). The odds ratios of the SNP combinations were found to be around 3.7 on average (Figure 4b).



(a) **(b)**

Figure 4: (a) Distribution of the combinatorial order of the 84 validated combinatorial disease signatures identified in the Pain Questionnaire cohort – i.e., 3 = signatures containing 3 co-associated SNPs. (b) Boxplot showing distribution of odds ratio and P1000 associated with 84 disease signatures identified in the Pain Questionnaire cohort.

Table 3 represents an example of a disease signature identified in this analysis containing five SNPs that were mapped to five genes.

Table 3: Example of one of the combinatorial disease signatures contributing to Community 1 identified by the PrecisionLife combinatorial analysis of the Pain Questionnaire cohort. Bold text indicates the critical (RF-scored) SNPs (and the genes to which they are mapped) in this signature.

SNP ID/Genotype	Mapped Genes	# Cases (t=2,382)	# Controls (t=4,764)
rs58634322 / 0 **rs7731560 / 2** rs77821266 / 0 **rs237475 / 1** **rs2904106 / 0**	**AC009486.1** AP001978.1 **KCNB1/** **AL035685.1** **ATP9A**	120	48

Patient Stratification

The disease architecture (Figure 5) generated by clustering[30] the SNPs in the disease signatures on the basis of patients in which they co-occur reveals the genetic heterogeneity of the ME/CFS Pain Questionnaire patient population, providing useful insights into patient stratification. These clusters ('communities') represent patient subgroups that (by definition) have shared disease etiology, and are therefore likely to share disease phenotypes, including severity, progression rate, clinical presentation, and, ultimately, therapy response.

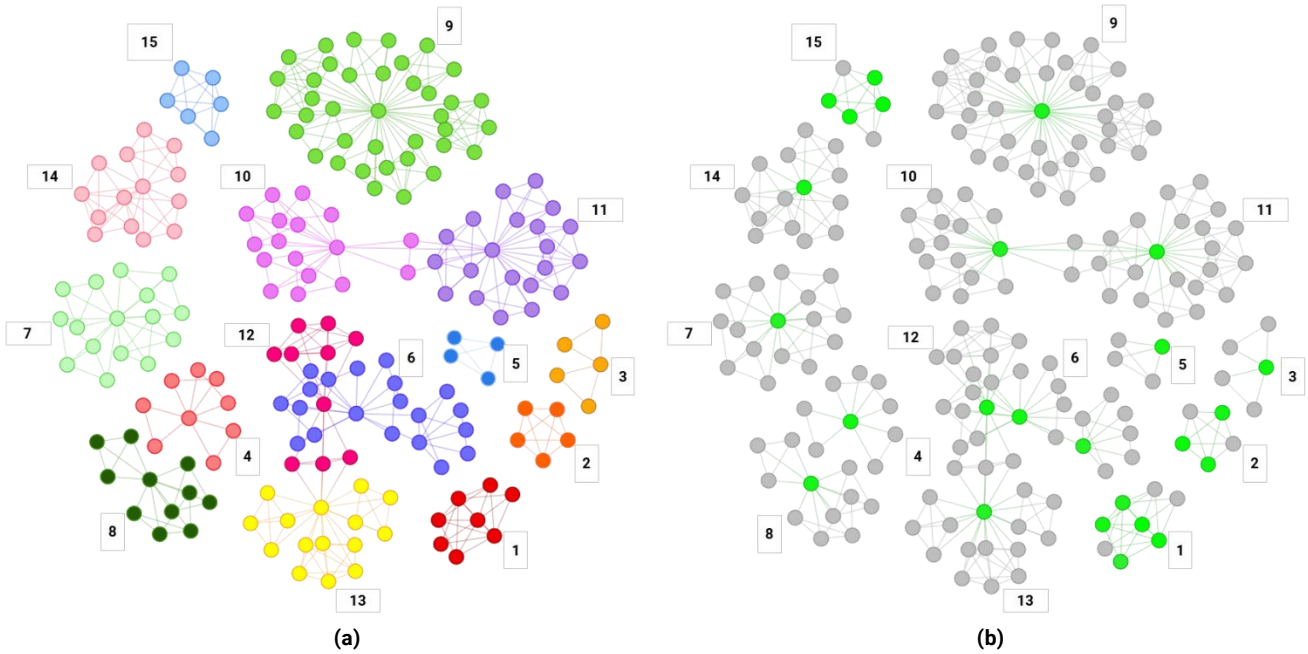


Figure 5: (a) Disease architecture diagram demonstrating the 15 communities of SNPs that make up the structure of the Pain Questionnaire patient sub-populations generated by the PrecisionLife platform. Each circle represents a disease-associated SNP genotype, edges represent their co-association in patients in disease signature(s), and colours represent distinct patient sub-populations. (b) the same disease architecture view coloured to show the critical SNPs associated with each community (light green)

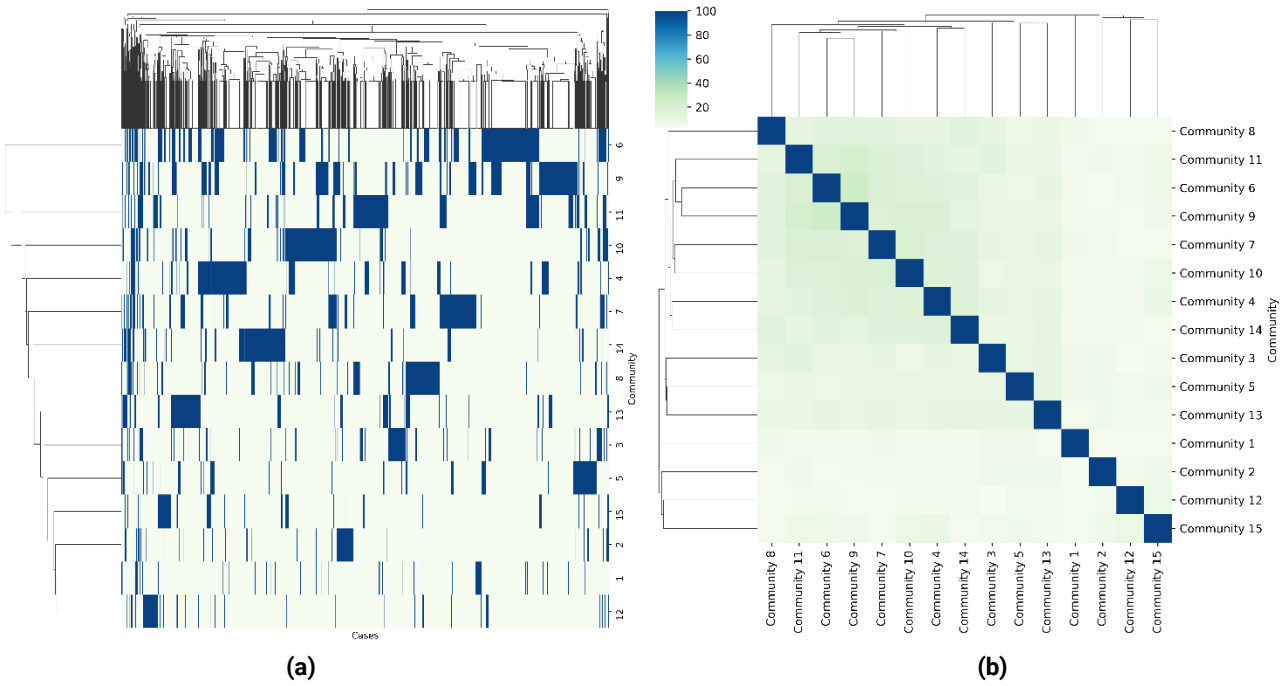


Figure 6: (a) Clustered heatmap showing the similarity of the 15 communities based on their respective cohorts of associated Pain Questionnaire patients. (**b**) Clustered heatmap showing the overlap of Pain Questionnaire patients associated with the 15 communities identified in the ME/CFS disease architecture.

There are 15 distinct communities of SNPs shown in the ME/CFS Pain Questionnaire disease architecture (Table 4). These share very low (<20%) patient overlap with each other (Figure 6), indicating they are distinct patient subgroups with different genetic drivers underlying their disease.

Table 4: Disease signatures, SNPs and case count associated with communities identified in the Pain Questionnaire cohort.

Community Number	Disease Signatures	SNPs	Case Count (t=2,382)	Control Count (t=4,764)	Chi Square *p*-value	Odds Ratio
1	6	8	144	102	2.56×10^{-17}	2.94
2	1	5	153	72	8.30×10^{-29}	4.47
3	2	5	257	203	5.13×10^{-26}	2.72
4	4	9	419	371	2.10×10^{-35}	2.53
5	2	4	228	197	8.46×10^{-20}	2.45
6	12	20	744	719	5.36×10^{-57}	2.56
7	7	16	408	315	1.18×10^{-43}	2.92
8	4	11	343	231	2.86×10^{-44}	3.30
9	17	38	648	503	1.60×10^{-72}	3.17
10	6	17	425	376	5.24×10^{-36}	2.53
11	6	21	469	348	5.55×10^{-54}	3.11
12	3	10	142	136	2.34×10^{-10}	2.16
13	6	15	282	207	5.08×10^{-32}	2.96
14	4	14	348	193	1.23×10^{-56}	4.05
15	2	6	170	89	5.94×10^{-29}	4.04

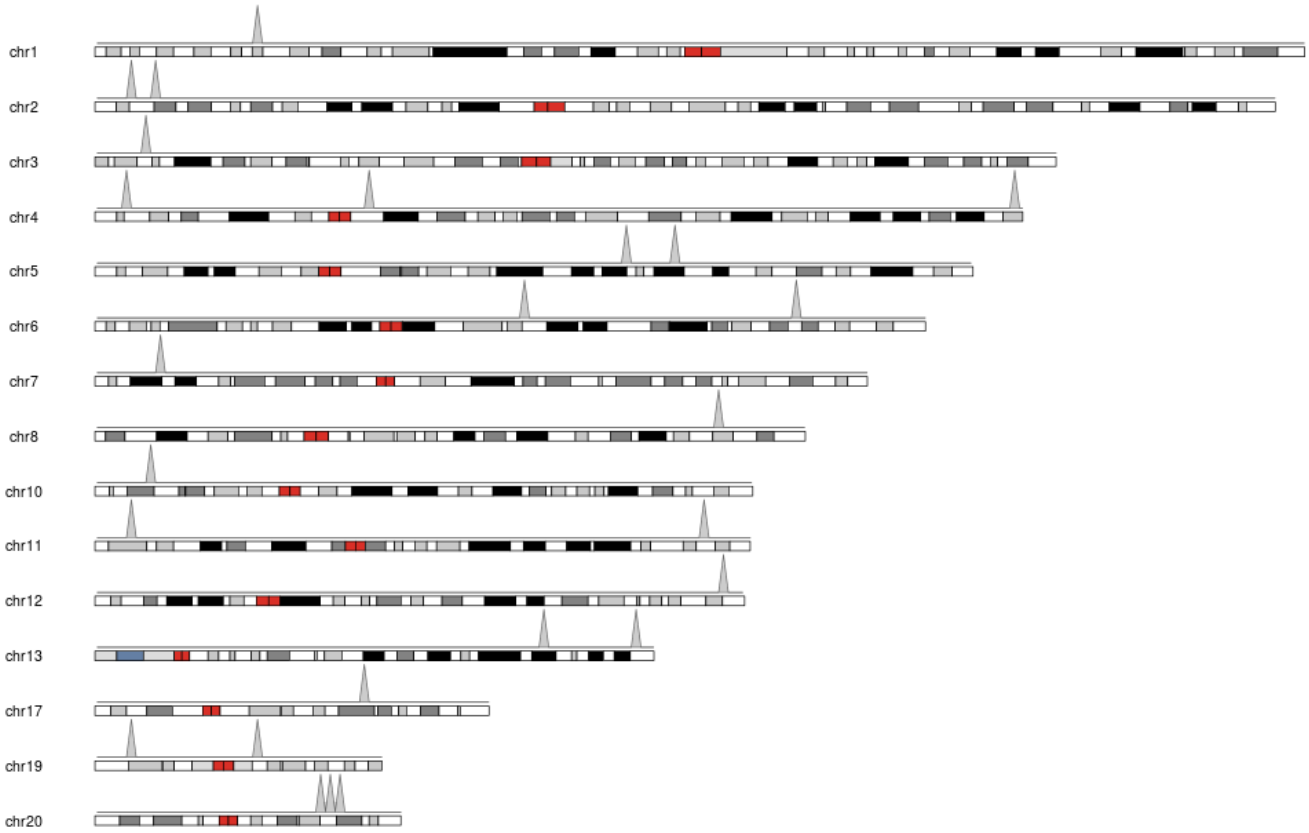


Figure 7: Autosomal locations for 25 critical disease associated SNPs identified in the Pain Questionnaire cohort.

The analysis identified 25 critical disease associated SNPs (see the Methods and Functional Genomics Annotation sections) which are identified in multiple disease signatures. These critical SNPs (Figure 7) were mapped to 14 protein coding genes strongly associated with the ME/CFS Pain Questionnaire case population (Table 5). Investigation of the function and mechanisms of action of these genes (and encoded proteins) revealed associations with one or more of five disease mechanisms that have been associated with ME/CFS development – viral/bacterial susceptibility, autoimmune development, metabolic dysfunction, vulnerability to stress, and sleep disturbance.

Enrichment analysis of available phenotypic and clinical data for the ME/CFS patients was used to generate additional insights into the clinical characteristics of each SNP community and prioritized gene. Statistical significance was calculated using the two proportions Z-test for categorical variables and the Mann-Whitney U test for continuous variables.

This analysis revealed 11 genes from 6 different patient communities with a level of enrichment with a particular phenotypic or clinical feature, such as increased incidence of clinical diagnosis of fibromyalgia or increased phenylalanine levels in plasma (Table 5), when compared against the rest of the case population. These associations, however, did not reach statistical significance ($p<0.05$) after multiple testing correction (Table 9 and Table 10 in Supplementary Data).

Table 5: Genes and communities identified in the Pain Questionnaire cohort associated with phenotypic and clinical features.

Gene	SNP (Most severe variant consequence – VEP)	SNP Minor Allele Frequency	Gene association with Phenotypic or Clinical feature (p-value before multiple testing correction)	Community	Community association with Phenotypic or Clinical feature (p-value before multiple testing correction)	Gene association replicated in disjoint Verbal Interview Cohort
S100PBP	rs41306603 (3 prime UTR variant)	0.052	Males (p=0.022)	1	Males (p=0.016)	No
ATP9A	rs2904106 (intron variant)	0.271	Males (p=0.016)	1	Males (p=0.016)	Yes
KCNB1	rs237475 (intron variant)	0.472	Males (p=0.016)	1	Males (p=0.016)	No
CLOCK	rs6832769 (3 prime UTR variant)	0.362	Fibromyalgia, ICD-10: M79.7 (p=0.026); Other soft tissue disorders, ICD-10: M79 (p=0.044)	2	Fibromyalgia, ICD-10: M79.7 (p=0.026); Other soft tissue disorders, ICD-10: M79 (p=0.044)	No
SLC15A4	rs2398428 (intergenic variant)	0.269		3		No
TMEM232	rs58264436 (intergenic variant)	0.394		6		No
GPC5	rs16947237 (intron variant)	0.155		7		No
PHACTR2	rs9403525 (intron variant)	0.448	Thyroiditis, ICD10-10: E06 (p=0.033)	8	Thyroiditis, ICD10-10: E06 (p=0.033)	No
AKAP1	rs3785477 (intron variant)	0.398		9		No
USP6NL	rs2499908 (intergenic variant)	0.235		12		No
CDON	rs73021223 (intron variant)	0.093	Illness, injury, bereavement, stress in last 2 years 2014+ (p=0.023)	13		No
INSR	rs59165976 (intron variant)	0.388		14		No

SLC6A11	rs2304725 (synonymous variant)	0.319	Phenylalanine levels in plasma (p=0.022); Illness, injury, bereavement, stress in last 2 years 2019+ (p=0.003)	15	Illness, injury, bereavement, stress in last 2 years 2019+ (p=0.003)	Yes
SULF2	rs56218501/ Affx-16805420 (missense variant)	0.213	Phenylalanine levels in plasma (p=0.022); Illness, injury, bereavement, stress in last 2 years 2019+ (p=0.003)	15	Illness, injury, bereavement, stress in last 2 years 2019+ (p=0.003)	No

Replication in Disjoint CFS (Verbal Interview) Cohort

We analyzed the disjoint UK Biobank CFS Verbal Interview cohort (1,273 cases, excluding individuals that were common to the Pain Questionnaire cohort, and 4,137 controls) using the PrecisionLife combinatorial analytics platform.

No SNPs were reported to be significant ($p<5 \times 10^{-8}$) for the Verbal Interview cohort in a standard GWAS case-control association analysis (Figure 8), however, the genomic loci showing modest association values around $p=1 \times 10^{-5}$ were found to be different than for the Pain Questionnaire cohort (Figure 3). This is likely due to the low power of the two GWAS, and could mean either that these are not true associations or, less likely, that the two populations simply yield different sets of true associations.

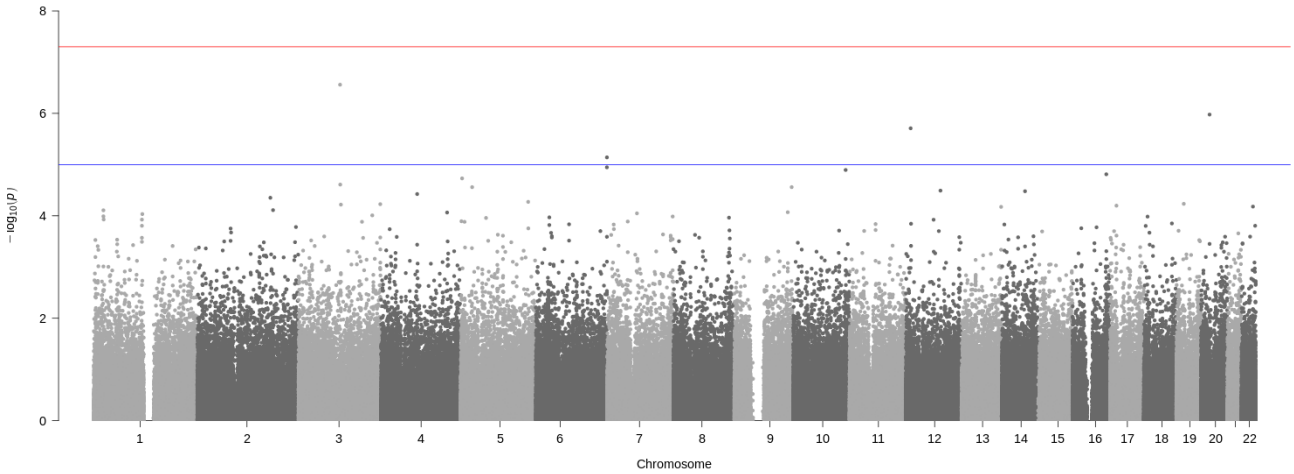


Figure 8: Manhattan plot generated using PLINK of genome-wide p-values of association for the disjoint Verbal Interview cohort (n = 5,140 where cases=1,273 and controls=4,137). The horizontal blue and red dashed lines represent the genome-wide significance values of $p<1 \times 10^{-5}$ and $p<5 \times 10^{-8}$ respectively.

Comparison of the results validating the 199 SNPs from the Pain Questionnaire cohort in the Verbal Interview cohort showed that five of the 25 critical SNPs (rs2304725, rs2904106, rs9444564, rs10420798, rs11695478) identified in the Pain Questionnaire cohort were replicated in this analysis. Two of these critical SNPs mapped to the *SLC6A11* (rs2304725) and *ATP9A* (rs2904106) genes, which were identified in this cohort. This suggests that these five critical SNPs (and two genes) are particularly strongly associated with ME/CFS. None of the replicated SNPs has significant direct GWAS associations to the disease or traits.

Disease Mechanisms and Genetic Functions

We used a detailed analysis of the metabolic context, exploiting an integrated semantic knowledge graph drawing from different data sources including Open Targets[110] associations, known gene-disease associations from scientific literature, mouse phenotypes etc., to annotate the 14 genes identified in the analysis of the Pain Questionnaire cohort.

While acknowledging annotation bias and an inevitable degree of subjectivity, we applied consistent heuristics to the available knowledge around a target, enabling us to identify that variants in these genes might impact different cellular processes. The five cellular processes or biological systems identified have previously been associated with ME/CFS– namely, susceptibility to infection, autoimmune and chronic inflammation development, metabolic dysfunction, increased vulnerability to stress and sleep disturbance – and it is possible to form plausible disease phenotype hypotheses for them (Table 6).

Table 6: Genes and communities identified in the Pain Questionnaire cohort and their proposed mechanism of action (MoA) in ME/CFS development (accounting for eQTL directionality).

Gene	Community	MoA Category	MoA Evidence
S100PBP	1	Viral/Bacterial Susceptibility	Identified as a differentially expressed gene (DEG) in response to RSV infection[31]
ATP9A	1	Metabolic	Exhibits P4-ATPase activity in beta cells and *ATP9A* knockdown results in reduction in glucose-stimulated insulin release[32]
		Autoimmune	Splicing region in *ATP9A* associated with multiple sclerosis risk in whole-exome sequencing analysis[33]
KCNB1	1	Metabolic	Regulates insulin exocytosis from pancreatic beta cells[34]
CLOCK	2	Sleep	Key regulator of circadian rhythm[35]
		Metabolic	Disruption to circadian clock results in mitochondrial dysfunction and insulin resistance[36,37]
SLC15A4	3	Autoimmune	SLC15A4 mediates type 1 interferon production and generation of autoantibodies in systemic lupus erythematosus (SLE)[38]
		Metabolic	Knock down of *SLC15A4* decreases mitochondrial membrane potential and impairs AMPK activation[39,40]
TMEM232	6	Autoimmune	The promoter region of *TMEM232* is differentially methylated in a study of multiple sclerosis patients[41]
GPC5	7	Autoimmune	Polymorphisms in GPC5 have been associated with multiple sclerosis risk in GWAS[42,43]
PHACTR2	8	Autoimmune	Identified as a multiple sclerosis risk gene in a GWA analysis[44]
AKAP1	9	Metabolic	AKAP1 is an AMPK substrate that regulates mitochondrial respiration[45]
		Viral/Bacterial Susceptibility	*AKAP1-/-* mice exhibit increased inflammation and immune cell infiltration in lungs[46]
USP6NL	12	Viral/Bacterial Susceptibility	Plays a role in HSV-1 virion assembly[47]
CDON	13	Metabolic	Cdon depletion results in impaired muscle regeneration and increased cell stress[48]
		Autoimmune	Cdon is involved in differentiation and myelination of oligodendrocytes[49]
		Viral/Bacterial Susceptibility	GWAS indicates *CDON* plays a role in complicated *Staphylococcus aureus* bacteremia[50]
INSR	14	Metabolic	Receptor tyrosine kinase which mediates insulin signalling[51]
SULF2	15	Viral/Bacterial Susceptibility	Genome-wide environmental interaction study indicates an association between a locus in *SULF2* and HSV-1 seropositivity and depression[52]
		Metabolic	Overexpression of *SULF2* observed in subjects with type 2 diabetes and obesity[53]
SLC6A11	15	Sleep	Associated with depression, sleep disturbance and neuropathic pain in mouse models[54,55]
		Autoimmune	Expression decreased in the hippocampus of mouse model of multiple sclerosis (EAE)[56]

Furthermore, critical SNPs found in the same patient community may be mapped to genes with shared biological functions or pathways. Pathway enrichment analysis of the genes using gprofiler[57] (excluding electronic Gene Ontology annotations), indicated that two large communities identified in this analysis containing multiple critical SNPs – Community 1 and Community 15 (Table 5) – may be implicated in common biological processes (Supplementary Table 14).

Community 1 contains three critical SNPs; rs41306603, a 3 prime UTR variant mapping to *S100PBP,* and two intronic variants, rs2904106 and rs237475, found in *ATP9A* and *KCNB1* respectively. The genetic variants in *ATP9A* and *KCNB1* were found in the same disease signature (an example is shown in Table 3), which shows significant enrichment linked to regulation of exocytosis and negative regulation of secretion by cell (GO annotations, Supplementary Table 14). Using additional evidence from the scientific literature, both *ATP9A* and *KCNB1* are expressed in pancreatic beta cells and are involved in the regulation of insulin secretion (Table 6)[36, 38]. This could suggest are combined biological effect of these two co-associated SNPs in causing dysregulated insulin signalling in this subgroup of ME/CFS patients.

Community 15 contains two critical SNPs; rs2304725, a synonymous variant in *SLC6A11*, and rs56218501/ Affx-16805420, a missense variant in *SULF2*. This disease signature shows enrichment for GABA synthesis and release and synaptic vesicle cycle (Reactome and KEGG annotations, Supplementary Table 14). This enrichment can be supported by further literature evidence that indicates the association of these genes with depression and other CNS-related disorders (Table 6)[56, 58, 59].

Many of the patient communities identified contain genes that could be categorized into more than one of these mechanisms and there was no clear distinction in biological pathways when communities were compared (Figure 9). This supports the hypothesis that development of ME/CFS is caused by the interaction and subsequent dysregulation of multiple immune, metabolic and neuronal pathways in combination.

We used the additional phenotypic and clinical data available in the UK Biobank to generate a patient profile for each patient community. However, the validation and significance of these findings are limited by the scope and depth of disease related data collected in UK Biobank (and other sources) that is available and relevant to ME/CFS patients, and the paucity of disease models.

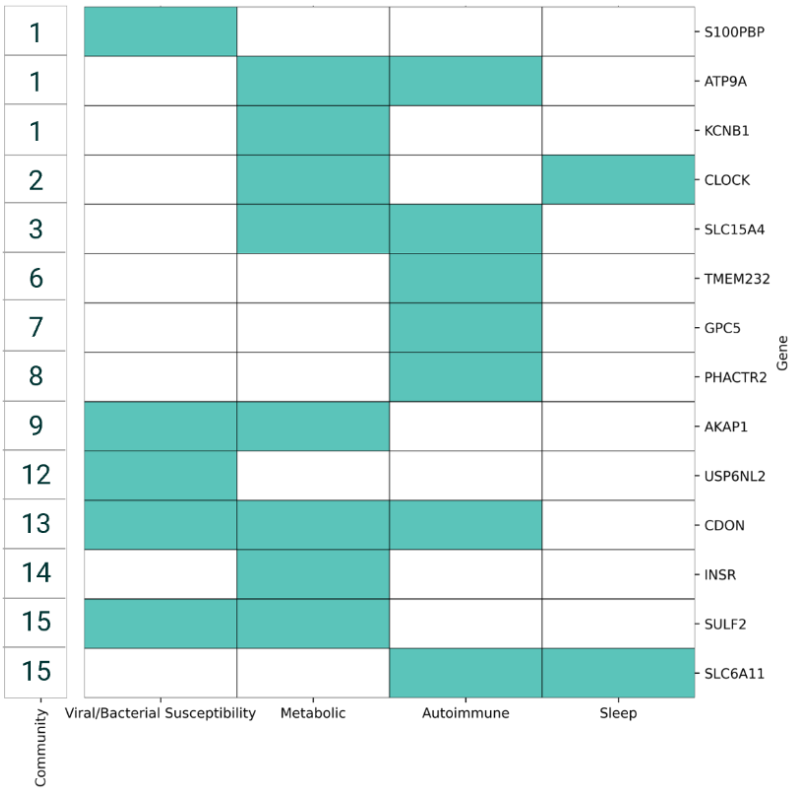


Figure 9: Biological pathways and processes known to be associated with the genes identified by the Pain Questionnaire study. Each border color represents a different patient community.

Viral/Bacterial Susceptibility

ME/CFS onset is often thought be linked to viral infection in patients, although no specific single viral or bacterial trigger has yet been confirmed[58]. There have been reports of shared pathophysiological, clinical, and transcriptomic features between viral and/or bacterial diseases and ME/CFS[59,60].

We identified five genes – *S100PBP, AKAP1, USP6NL CDON* and *SULF2* – in five different patient subgroups that have been associated with viral and/or bacterial infection in the literature (Table 6).

These may represent a subset of ME/CFS patients with increased susceptibility to infection, or differential response to infection that leads to ineffectual viral clearance. We therefore evaluated the clinical records of the ME/CFS case population included in this study to identify any evidence of prior infection of the most common ME/CFS-associated infective triggers, including infectious mononucleosis, and EBV and/or Herpesviruses seropositivity. Unfortunately, the total numbers of patients and clinical reports with any of these was too small (approximately 2% of cases) to generate any statistically significant gene/patient subgroup associations, so the question of whether any such significant associations exist remains unanswered.

Autoimmune and Chronic Inflammation

Our analysis identified seven genes that have been associated with diseases that have autoimmune components in both the literature and in other disease studies that we have undertaken, including COVID-19, rheumatoid arthritis and Sjögren's syndrome (unpublished results).

ME/CFS shares several characteristics with autoimmune diseases, including the increased level of pro-inflammatory cytokines and higher prevalence in females, with as many as 60% of ME/CFS patients also reported to be diagnosed with an autoimmune disease[61,62].This co-association with other autoimmune diseases was also evident in our analysis of ME/CFS patients when compared against the rest of the UK Biobank population (Figure 2). Whether this reflects a real association or misdiagnosis of patients remains unclear.

We speculate that increased susceptibility to viral infection in ME/CFS patients, resulting in recurrent or chronic infections, may also drive chronic inflammation and autoimmune development.[63] Furthermore, pro-inflammatory cytokines associated with autoimmune development have also been shown to contribute to mitochondrial dysfunction and decreased respiratory capacity, and there is evidence that patients with other autoimmune diseases also display mitochondrial dysfunction[64,65,66].

Solute carrier family member 15 (SLC15A4) is found in the lysosomal membrane and has enriched expression in immune cells. Genetic variants in *SLC15A4* have been associated with increased risk of developing inflammatory diseases like systemic lupus erythematosus.[67] Interestingly, SLC15A4 has been shown to play a crucial role in immune cell tolerance to metabolic stress via AMPK and mTORC1 and maintenance of respiratory homeostasis in innate immune cells[68] and *SLC15A4* knock down results in decreased mitochondrial function under cell stress[69]. No eQTL associations were found for the ME/CFS SNP linked to *SLC15A4*.

A specific variant – associated to *GPC5* (glypican 5) – was found in 17% (408) of ME/CFS cases in the Pain Questionnaire study. Glypican 5 is a cell surface proteoglycan that has been identified in many different multiple sclerosis genetic studies[70,71,72]. A further four – *ATP9A, TMEM232, PHACTR2* and *SLC6A11* - out of the seven autoimmune genes identified in this study can also be linked to multiple sclerosis development (Table 6). MS and ME/CFS are believed to have a viral trigger component, such as Epstein-Barr virus (EBV), and their patients share similar symptoms, including fatigue, pain, sleep disturbance and cognitive dysfunction[73].

Metabolic Dysfunction

Reductions in reserve capacity and inability to raise mitochondrial respiration in response to stress compared with controls indicates that ME/CFS patients are less able to meet energy demands, resulting in increased fatigue and exercise intolerance[74].

Combinations of genes including a variant in *AKAP1* were found in 27% (648) of the Pain Questionnaire cases – the highest proportion for any RF-scored genetic variant identified in the study – and no more than 1.5% of controls. AKAP1 (A kinase (PRKA) anchor protein 1) is a scaffold protein in the mitochondrial membrane, regulating mitochondrial respiration via AMPK. A study has demonstrated that phosphorylation of AKAP1 by AMPK was crucial for AMPK-induced increase in mitochondrial respiration in human muscle post-exercise.[75] Furthermore, knockout of *AKAP1* in mice resulted in reduced skeletal muscle capillary density and functional recovery impairment in addition to increased mitochondrial dysfunction and cellular stress in endothelial cells[76]. The identification of a disease associated *AKAP1* variant in this study provides a strong genetic link to mitochondrial dysfunction and the reduction in energy capacity observed in biochemical analysis of ME/CFS patients[77].

We also identified a series of genes involved in other metabolic processes such as insulin sensitivity and lipid metabolism.

ATP9A is a member of the Type IV P-type ATPases (P4-ATPases) family involved in the process of lipid flipping. ATP9A may regulate intracellular levels of ceramide and sphingosine[78], which have been shown to be altered in patients with chronic fatigue and in the skeletal muscle of fatigue-associated conditions[79,80,81]. ATP9A is also expressed in pancreatic beta cells and has a role in driving glucose-stimulated insulin release[82]. Moreover, a variant in *ATP9A* has been associated with multiple sclerosis in a homozygosity haplotype analysis[83].

Finally, 348 (15%) ME/CFS patients (and 4% of controls) from this study were most associated with the community of genetic variants including the gene encoding the insulin receptor (*INSR*). A study has found that insulin levels in ME/CFS patients were higher than in healthy controls[84], which is hypothesized to be as a results of insulin resistance and ischemia-reperfusion damage in skeletal muscles of patients with ME/CFS[85].

These 348 ME/CFS patients also presented with relatively higher blood levels of lactate from the UK Biobank NMR metabolomics data ($p<0.017$, Supplementary Table 11) compared to the entire case population. Lactate is implicated in insulin resistance, resulting in reduced insulin-dependent glucose uptake in skeletal muscle and dysregulated insulin signaling[86,87]. Dysregulation of insulin and lactate in ME/CFS patients may also have an impact on mitochondrial function, decreasing mitochondrial size and respiratory function[88].

Response to Stress

Three of the genetic variants that were significant in the Pain Questionnaire analysis – located in genes *SLC6A11*, *SULF2* and *CDON* – were identified in communities of ME/CFS patients more likely ($p=0.003$, Supplementary Table 10) to report the occurrence of illness and psychosocial factors (injury, bereavement, stress) in the last 3-8 years. These could represent a subset of ME/CFS patients with combinations of variants involving these genes that confer vulnerability to psychological stress.

SLC6A11 (GAT3) is a sodium-dependent transporter involved in GABA reuptake at presynaptic terminals. Altered levels of GAT3 have been associated with increased neuroinflammation and cognitive impairment[89], in addition to sleep disturbance, juvenile stress and depression in animal models[90,91,92,93]. Furthermore, patients with SNP combinations including those in *SLC6A11* display increased levels of phenylalanine ($p=0.022$) in the metabolomics data compared to other ME/CFS subgroups identified in our analysis. Phenylalanine is a precursor for monoamine neurotransmitters, such as dopamine, epinephrine and serotonin. Finally, two further SNPs in *SLC6A11* were also identified to be significant in the Verbal Interview ME/CFS case dataset, providing additional evidence for the importance of this gene in ME/CFS development.

Sulfatase 2 (SULF2) is an enzyme that regulates the effects of heparan sulfate. A variant in this gene is found in combinations with *SLC6A11* and is therefore also associated with the patient community with raised phenylalanine levels. Sulfatase 2 plays a role in a wide variety of biological process and is expressed in most tissues (Figure 16 in Supplementary Data). Sulfatase 2 is crucial for brain development, contributing to processes such as neurite outgrowth and responsiveness to growth factors[94,95,96], and there is an association between *SULF2* variants and HSV-1 and depression risk, and also with malaise and fatigue in UK Biobank studies[97,98]. Although there is a known association with sex hormone globulin levels, we did not find any difference in male:female distribution for this community.

CDON (cell adhesion associated, oncogene regulated) encodes a cell surface receptor that is highly involved in muscle regeneration[99]. In muscle cells, depletion of CDON results in impaired muscle regeneration and senescence, as well as increased cell stress[100]. However, CDON has also been associated with complicated bacteremia[101] and development of midbrain dopamine pathways[102]. This indicates that CDON has a diverse range of functional roles that could impact ME/CFS development.

Sleep Disturbance

We identified two genes that could play a role in the sleep disturbance often reported by ME/CFS patients, *SLC6A11* and *CLOCK*. The *CLOCK* (Circadian Locomotor Output Cycles Kaput) gene is one of the key regulators of circadian rhythm. Altered circadian rhythm is hypothesized to contribute to many of the symptoms experienced by patients with ME/CFS, including insomnia, pain and post-exertional malaise.[103] This is because disruptions in the circadian clock have far reaching biological consequences beyond sleep disruption, including disturbed mitochondrial function, dysregulated cellular stress responses and insulin sensitivity[104,105,106]. Furthermore, transcriptomic analysis of peripheral blood mononuclear cells indicated that several genes involved in circadian rhythm were elevated in ME/CFS patients[107].

We also found significant enrichment in patients with variants in *CLOCK* who also had been diagnosed with fibromyalgia. ME/CFS and fibromyalgia patients exhibit similar symptoms, including fatigue, cognitive functioning impairment and pain, which could indicate similar underlying biological drivers of disease (or a degree of misdiagnosis). Interestingly, a study investigating the differences between the two conditions found that patients with both ME/CFS and fibromyalgia also presented with sleep disruption, in contrast to CFS only patients and healthy controls[108]. These results could reveal further insights into the cause of this symptom.

Drug Target Evaluation

There are currently no specific pharmacological treatment options for ME/CFS patients. The detailed insights generated by combinatorial analysis of this UK Biobank population can be used to inform the development of novel drug targets guided by patient stratification biomarkers associated with each of the ME/CFS subgroups.

In other studies, for example in motor neuron disease / amyotrophic lateral sclerosis (ALS), we have at this stage identified known pharmacological modulators of several novel targets discovered using the approach described above. We tested these in a patient-derived human induced neuronal progenitor cells (iNPC) cellular assay with a co-culture of motor neurons, microglia and astrocytes[109] to provide biological validation of the disease modification potential of modulating several novel targets identified using this methodology (manuscript in preparation). We are further developing direct CRISPR derived knock-in/knock-outs for those targets in iPSC-derived neurons. However, in ME/CFS, not only are there no assays or model systems available, but we also do not understand the tissues involved in various aspects of the disease. This prevents the ready evaluation of the effects of modulating the targets either pharmacologically or via direct genetic manipulation.

Each gene identified in this study was evaluated for drug tractability (Table 7), indicating that seven of the targets exhibit potential small molecule or antibody tractability. Moreover, three of the genes are targeted by drugs in clinical development, suggesting their potential as drug repositioning candidates, which might offer a faster and derisked route to approval if their safety and efficacy can be demonstrated.

Table 7: Genes identified in the Pain Questionnaire cohort with their tractability as drug targets using annotations from OpenTargets[110].

Gene	Small Molecule Tractable	Small Molecule Clinical	Antibody Tractable	Existing Drugs	Development Compounds	PDB Structures
S100PBP	0	0	0	0	0	0
ATP9A	0	0	1	0	0	0
KCNB1	0.3	1	1	7	28	0
CLOCK	0	0	0	0	0	2
SLC15A4	0	0	1	0	0	0
TMEM232	0	0	0	0	0	0
GPC5	0	0	0.3	0	0	0
PHACTR2	0	0	0.3	0	0	0
AKAP1	0	0	0	0	0	0
USP6NL	0	0	0.3	0	0	0
CDON	0	0	0.3	0	0	3
INSR	1	0.7	0.3	20	891	62
SULF2	0	0	0	0	0	0
SLC6A11	0.3	0	0.3	6	126	0

Discussion

After decades of study, the genetic contributions to the etiology of ME/CFS and the different mechanisms underpinning the disease remain poorly understood. It is unsurprising therefore that our analysis demonstrates that ME/CFS at a genetic level is polygenic and heterogeneous.

This is confirmed both by the genetic association and patient stratification results generated using combinatorial analysis techniques in this study, as well as the consistent failure of previous GWAS analyses to find replicable signal within this cohort and/or between ME/CFS population datasets, which would be expected if clinically relevant monogenic signals were present[6].

Using a hypothesis-free combinatorial analytics approach based on the PrecisionLife platform, we identified 199 SNPs in 84 high-order combinations that were highly associated with 91% of the ME/CFS cases in the UK Biobank Pain Questionnaire cohort. These variants could be mapped to 14 genes, which appear to be compatible with the major cellular mechanisms suspected by other groups working in the field and show a level of overlap with diseases sharing similar symptoms, such as MS[111]and long Covid[112].

We further used these findings to stratify the ME/CFS patients genetically and correlated this stratification with clinical criteria. There is a degree of evidence of replication of several SNPs and two of those genes being identified in a second UK Biobank cohort, and the consistency of results from internal cross-validation replication runs is also encouraging.

Biological analysis of these genes indicates that many of them are directly linked to the key cellular mechanisms hypothesized to underpin ME/CFS, including vulnerabilities to stress and infection, mitochondrial dysfunction, sleep disturbance and autoimmune development. This has revealed several potential novel drug targets that could be the basis of targeted therapy development for ME/CFS patients.

Study Limitations

There are however a number of limitations with this study. Analysis of ME/CFS data is complicated by several logistical factors impacting data availability and quality, including low reporting rates, inaccurate diagnosis, limited cohorts with genetic information, and limited longitudinal clinical, psychosocial, epidemiological, and environmental data. This is exacerbated by the nature of the disease with its complex interactions of multiple etiologies, mechanisms, and influences.

The UK Biobank cohort, while essential to enabling this analysis, represents only a small cohort of atypically older ME/CFS patients with predominantly white, European ancestry who have self-reported their clinical diagnosis. The lack of detailed ME/CFS-specific supporting clinical and/or phenotypic data makes it hard to evaluate individual clinical experiences and assess potential triggers of disease onset or relapse.

While we have tried to replicate the analysis and results between two different UK Biobank cohorts, a high rate of false negatives, the self-reporting of the clinical diagnosis, which in some cases may be misdiagnosed, and other variations in the case criteria between the cohorts make expectation of a complete correlation of results unrealistic. It is nonetheless encouraging that five critical SNPs and two of the genes identified do in fact appear in both cohorts, even allowing for the shared genetic ancestries of the cohorts.

Although it can occur at any time of life, the average age at onset of ME/CFS is in the 30s[113], perhaps with an earlier secondary peak[114], whereas the average age of the UK Biobank population is 56 years[115] and the population has a selective participation bias to 'healthy volunteers'[116]. In the Pain Questionnaire study, the average age of cases was 69 years, indicating an even greater bias to a more elderly population. This might cause the associations identified to be skewed away from causes that could be more prevalent in a more age inclusive population or towards comorbidities that exerted a larger influence. On the other hand, an older population may be more accurately diagnosed. A better distribution of ages and longitudinal follow-up data would enable analysis of differences in etiology, clinical presentation or comorbidities and prescriptions.

ME/CFS is clearly a complex disease with multiple endogenous and exogenous triggers, potentially ranging from metabolism, autoimmune and infection, to stress and environmental impacts. Not all of these factors are recorded consistently and accurately in the available dataset, making their influence across one of more of the patient subgroups hard to determine definitively.

Finally, there is a considerable bias in the makeup of the patients both in UK Biobank and in this study. All of the participants in this study have a European ancestry due to their predominance in the source data[22]. There may well be different and additional mechanisms influencing the disease in cohorts with other ancestries and geographies (including different triggering pathogens).

Similarities with other Diseases

MS and ME/CFS patients share a number of similar symptoms, including pain, sleep disturbance and cognitive dysfunction[117], and both can have a viral trigger such as Epstein-Barr virus (EBV)[4,118]. There is also increasing evidence that many patients diagnosed with long COVID share similar symptoms, such as chronic fatigue and 'brain fog', with individuals with ME/CFS. It is also believed that some patients may be developing ME/CFS as a direct result of having a COVID-19 infection[119,120,121].

This suggests that the two diseases may share similar etiologies with possible overlap in the biological drivers and risk genes. Our analysis of the first UK Biobank COVID-19 population identified four genes out of 68 associated specifically with the risk of severe COVID that we had previously identified as having strong association with neurodegenerative processes[23], including *ATXN1, SORCS2* and *STH* and *MAPT* from loci on chromosome 17 that were subsequently validated by the results from the COVID-19 Host Genetics Initiative[122]. This analysis also revealed several other disease and symptom associated mechanisms, such as viral host response factors and pro-inflammatory cytokine production.

We are in the process of analyzing two populations in long COVID-19 (Sano Genetics, GOLD study) and multiple sclerosis (UK Biobank) in order to identify any shared genes and biological mechanisms underpinning ME/CFS, multiple sclerosis and long COVID-19 development. Preliminary findings from our long COVID analysis have indicated that three of the genes identified in this study are also significant in the long COVID patient group (albeit with different SNPs, but again none of these are in LD). These will be subject of further validation in a new publication later this year.

Conclusion/Future Perspectives

The use of a hypothesis-free combinatorial analytics approach using the PrecisionLife platform has enabled us to identify 14 novel genetic associations with ME/CFS in a UK Biobank cohort. Several previous attempts at GWAS approaches[12] have failed to validate single SNP associations or highlight significant risk genes in this ME/CFS cohort.

This study has produced further evidence of the polygenic and heterogeneous nature of the disease and produced patient stratification results that describe the mechanistic etiology of the disease. This also suggests a set of novel potential drug targets that may be relevant for the major ME/CFS patient subgroups.

There are a number of limitations with this study discussed above, and a larger, more detailed longitudinal patient dataset is likely to significantly improve the results. For this reason, we aim to replicate and extend the results from this UK Biobank study with combinatorial analysis of a future DecodeME study. DecodeME is the largest current genetic ME/CFS study, with over 20,000 participants involved[123], and the more detailed patient survey data collected is likely to allow deeper insights into the different subgroups and targets involved with the disease.

The findings of this study nonetheless provide some indicators of useful areas of study in terms of diagnostics, novel drug targets, and potentially precision repositioning opportunities. As a first step, simply identifying and validating patient stratification biomarkers that could be used to create an accurate risk model or diagnostic test for ME/CFS would be a huge step forward in recognition and treatment of the disease.

Discovery of drug candidates for ME/CFS has been limited in progress not just due to lack of plausible targets (and disease involved tissues), but also access to accurate models of the various aspects of the disease. Biological validation of the disease modification potential of the identified targets *in vitro* or *in vivo* is the next obvious step, but the lack of ready access to validated assays and disease models, or even a specific cell type to target is a barrier.

We hope that with a smaller set of genes on which to focus, genetic interventions (e.g., CRISPR knock in/out) or transient siRNA modulation might enable us to generate cell lines that capture features of the disease biology and to investigate in a cellular system the role that each target gene plays. We could further use these modified/modulated cell lines as assays to evaluate recovery of a normal phenotype in the presence of active molecules to accelerate the discovery and validation of novel and/or precision repositioned therapeutics.

We have identified known active compounds acting at three of the targets found in this study using precision repositioning approaches[124], and there is the potential to evaluate the likely impact of these retrospectively via analysis of real-world data collections with longitudinal prescription information, and also pharmacologically in the new assay systems using known active drugs and/or development candidates as tool compounds. Given a good safety profile for these compounds or their derivatives, this may provide sufficient evidence in the future for the design of first in man studies.

Finally, understanding the drivers of ME/CFS and disorders with similar symptoms such as long COVID and MS, and establishing the similarities and differences between them in more detail is likely to have profound implications for patients. Accurate diagnosis and effective treatment options are limited in all of these diseases, and we hope that uncovering of the disease etiologies, better patient stratification, and identification of novel drug targets will yield rapid progress in approval of better diagnostic tools and drugs for patients.

Declarations

Ethics Approval

Research described in this article has been conducted using data from UK Biobank Resource (application number 44288). UK Biobank has approval from the North West Multi-centre Research Ethics Committee (MREC) as a Research Tissue Bank (RTB) approval, and researchers do not require separate ethical clearance and can operate under this RTB approval.

Consent for Publication

Not applicable

Data Availability

All data sources are described in the Supplementary Information, and no new source data were collected. Only data from existing UK Biobank study cohorts were analyzed. All datasets generated during the study are described in the Supplementary Data section and/or available from the corresponding author upon reasonable request.

Competing Interests

S.D., K.T., J.K, J.S, and S.G. are employees of PrecisionLife, Ltd. S.G. is a shareholder of PrecisionLife, Ltd.

Funding

The project was funded entirely by PrecisionLife Ltd.

Authors' contributions

S.G., S.D., and K.T. wrote the manuscript. S.G. designed the approach, and K.T., S.D., J.K., and J.S. analyzed the data generated by the PrecisionLife platform. All authors provided input and approved the final version of the manuscript.

Acknowledgements

Research described in this article has been conducted using data from UK Biobank Resource (application number 44288). We would like to acknowledge the helpful advice and encouragement provided by Prof. Chris Ponting, MRC Investigator at the MRC Human Genetics Unit, Institute of Genetics and Cancer of the University of Edinburgh and Sonya Chowdhury, CEO of Action for ME. Special thanks to Matthew Pearson, Karan Dahele, and Mark Strivens who provided input into the manuscript, Gert Møller, who initially developed the combinatorial analytics methodology, and the rest of the PrecisionLife team.

Supplementary Data

Pain Questionnaire Study Design

UK Biobank participants included in the Pain Questionnaire case cohort answered positively to the following criterion:

Data Type	ICD-10 / Data Field ID	Description
Self-reported	120010	Ever had Chronic Fatigue Syndrome or Myalgic Encephalitis (M.E.)

UK Biobank participants excluded from the Pain Questionnaire control cohort met one or more of the following criteria:

Table 8: Control selection criteria for the ME/CFS Pain Questionnaire cohort

Data Type	ICD-10 / Data Field ID	Description
HES	G93.3	Postviral fatigue syndrome
HES	R53.x	Malaise and fatigue
HES	G47.x	Sleep disorders
HES	M79.7	Fibromyalgia
Self-reported	120009	Ever had fibromyalgia syndrome?
Self-reported	120010	Ever had Chronic Fatigue Syndrome or Myalgic Encephalitis?
Self-reported	120120	Exercise brings on fatigue
Self-reported	120122	Fatigue interferes with physical functioning
Self-reported	120123	Fatigue causes frequent problems
Self-reported	120124	Fatigue prevents sustained physical functioning
Self-reported	120125	Fatigue interferes with carrying out certain duties or responsibilities
Self-reported	120126	Fatigue is among three most disabling symptoms
Self-reported	120127	Fatigue interferes with work, family or social life

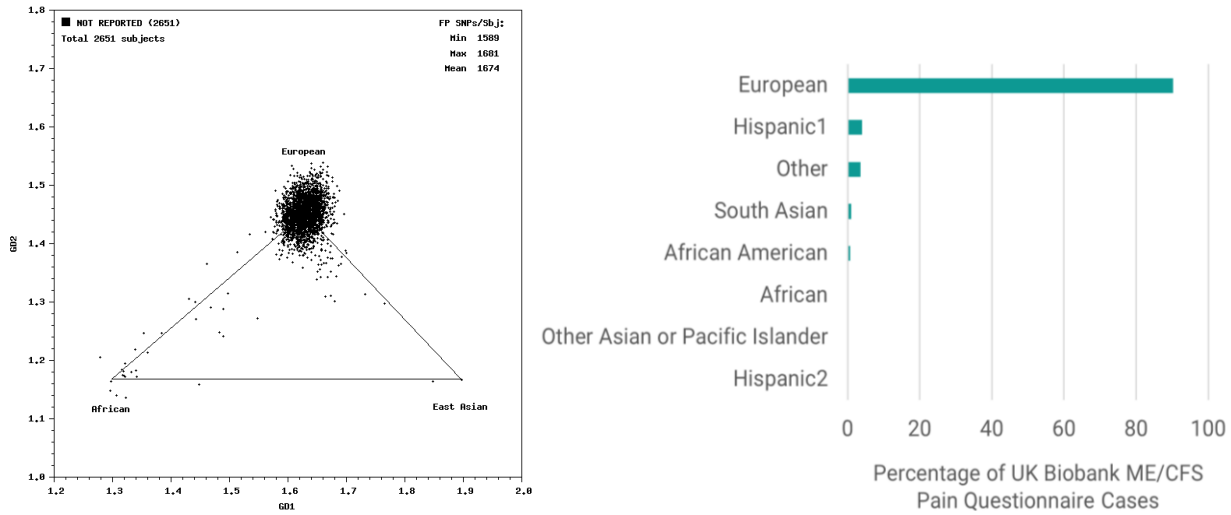
Genotype Quality Control

Appropriate quality control of genotype data was performed using PLINK[29] and GRAF[125] (Genetic Relationship and Fingerprinting) based on standard quality control procedures to ensure thorough cleaning of the data before it is used for genomic analyses.

This included the following steps:

1. Batch effect correction: Batch-level quality control procedures was performed based on recommendations by UK Biobank[22] and only SNPs that pass all batch-level QC were used for further analysis.
2. Sample and SNP filtering based on missingness: The filtering for SNPs with missing data (<5%) was followed by filtering of individuals with missing data (<5%) using PLINK.
3. Minor Allele Frequency (MAF) filtering of SNPs: The genotype data would be filtered to exclude SNPs with MAF <0.0001 using PLINK.
4. Hardy-Weinberg Equilibrium (HWE) filtering: HWE filtering was performed on controls with $p<10^{-10}$ using PLINK.
5. Heterozygosity filtering: Samples with extreme (very high or very low) heterozygosity were removed.
6. Sample filtering based on relatedness: GRAF-rel[126] was used to identify duplicates and closely related subjects in the dataset. After identification of close relatives, only one representative of each closely related family pairs was retained.
7. Ancestry analysis: GRAF-pop was used for ancestry inference and limit samples for the dataset to European ancestry.

8. Sex discrepancy of individuals: Samples that have discrepancies between the sex recorded in the dataset and their sex based on absence/presence of a Y chromosome were removed.



(a) **(b)**

Figure 10: (a) Ancestry inference plot generated by GRAF-pop and (b) the ancestry distribution of ME/CFS case population (n=2,651 cases) generated from UK-Biobank using self-reported diagnosis in Pain Questionnaire before quality control shows very strong bias for European ancestry (>90%).

Cohort Analysis

Data used for Cohort Analysis

Data used for cohort analysis in Figure 2:

1. Exposure to infectious agents:

Data Type	ICD-10/ Data Field /Category ID	Description
Blood assays	23053	EBV seropositivity for Epstein-Barr Virus
Blood assays	1307	Infectious disease antigens
HES	A00.x-B99.x	Chapter I Certain infectious and parasitic diseases

2. Diagnosis with any of the most common autoimmune diseases:

Data Type	ICD-10 / Data Field / Category ID	Description
HES	M05.x or M06.x	Rheumatoid arthritis
HES	L40.x or M07.x	Psoriasis / psoriatic arthritis
HES	G35.x	Multiple sclerosis
HES	M32.x	Systemic lupus erythematosus
HES	K50.x or K51.x	Inflammatory bowel disease (Crohn's and ulcerative colitis)
HES	K90.0	Celiac disease
HES	M45.x	Ankylosing spondylitis
HES	M35.0	Sjogren's syndrome
HES	M35.3	Polymyalgia rheumatica
HES	M34.x	Systemic sclerosis
HES	M79.7	Fibromyalgia
HES	E10.x	Type 1 diabetes
HES	E06.x	Thyroiditis

3. Evidence of significant stressful events:

Data Type	ICD-10 / Data Field /Category ID	Description
Self-reported	6145	Illness, injury, bereavement, stress in last two years (any but 'none of the above')
Self-reported	20499	Ever sought or received professional help for mental distress
Self-reported	20500	Ever suffered mental distress preventing usual activities
HES	Chapter XX	External causes of morbidity and mortality

Sex

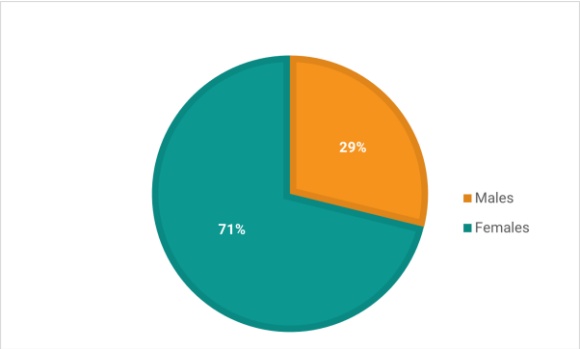
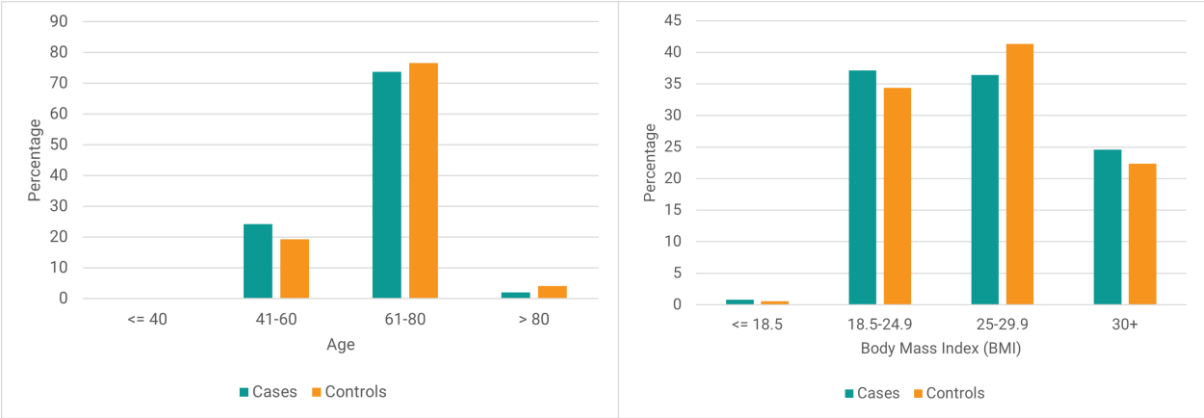


Figure 11: The proportion of females in the Pain Questionnaire case population was substantially higher (~71%) than males (29%).

Age & BMI



(a) **(b)**

Figure 12: Distribution of (a) age and (b) BMI of cases vs controls in the Pain Questionnaire cohort.

Combinatorial Disease Signatures

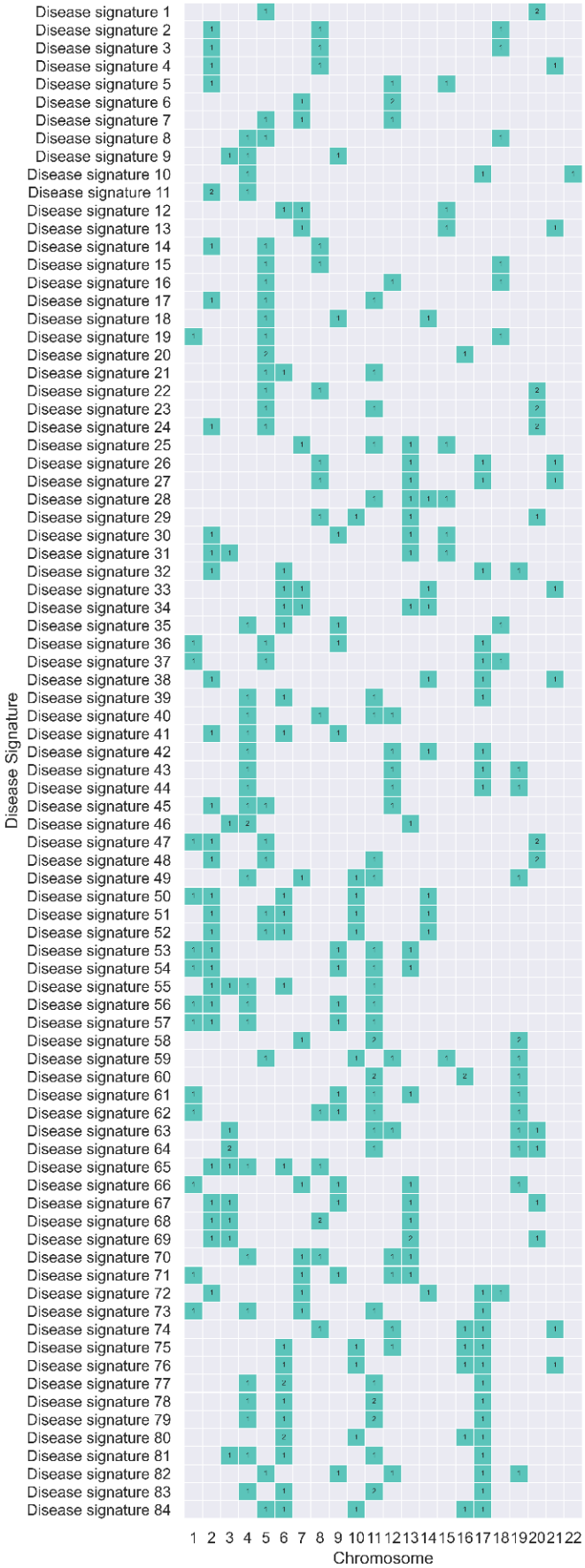


Figure 13: Distribution of chromosomal location of SNPs associated with 84 disease signatures identified in the Pain Questionnaire study. None of the SNPs identified in the disease signatures were observed to be in linkage disequilibrium (LD).

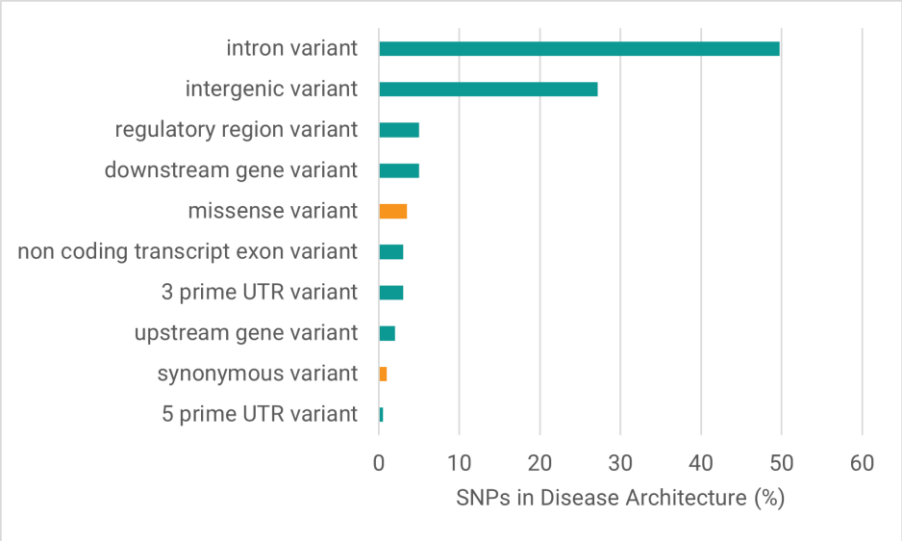
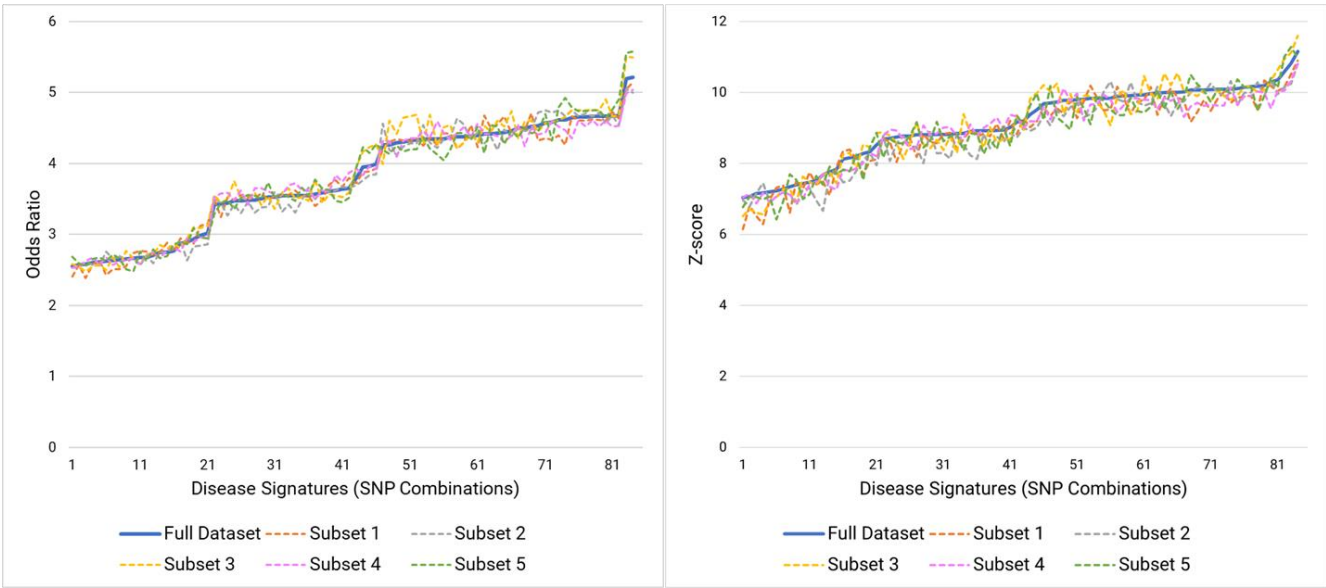


Figure 14: Distribution of variant consequences (most severe predicted by Ensembl VEP) of critical SNPs identified in the Pain Questionnaire study. More than 95% of SNPs were non-coding variants (shown in green) and <5% were coding variants (shown in orange).

As an internal validation we generated five smaller subsets of the Pain Questionnaire cohort each excluding a different 10% of the cases. These cross-validation runs therefore comprised 90% of the case population compared against the same set of controls as the Pain Questionnaire cohort. We compared the odds ratios and Z-scores of the disease signatures identified in the full cohort to those identified in each of the subset analyses. The odds ratios and Z-scores of the SNP combinations remain largely consistent irrespective of small changes to the case population (Figure).

This technical (internal cross-validation) replicate study provided an internal parameter check, but it has insufficient independence between both the case and control sets used in the runs for reliable use as formal replication.



(a) **(b)**

Figure 15: Comparison of (a) odds ratios and (b) Z-scores of the full Pain Questionnaire dataset with five sub-cohort splits containing 90% of the same cases as the full cohort and the same set of controls.

Genes Associated with Phenotypes

Table 9: Frequency distribution of categorical phenotypic features in cases associated with gene(s) compared to all cases and controls in the cohort. *p*-values were calculated to assess the association of each feature using two-sided Fisher's exact tests.

Phenotypic Feature	Gene(s)	Gene case count with feature	Gene Case count	All case count with feature	Total case count	All controls count with feature	Total control count	*p*-value (compared to cases) before multiple testing	Adjusted *p*-value (compared to cases)	*p*-value (compared to controls) before multiple testing	Adjusted *p*-value (compared to controls)
Illness injury bereavement stress in last 2 years (2019+)	*SLC6A11 SULF2*	4	170	12	2,382	1	4,764	0.003	0.177	6.00×10^{-21}	7.61×10^{-21}
Illness injury bereavement stress in last 2 years (2014+)	*CDON*	48	328	249	2,382	3	4,764	0.023	0.861	6.36×10^{-145}	1.19×10^{-144}
Fibromyalgia (ICD-10: M79.7)	*CLOCK*	13	153	108	2,382	0	4,764	0.026	0.969	2.92×10^{-90}	1.87×10^{-89}
Other soft tissue disorders (ICD-10: M79)	*CLOCK*	22	153	224	2,382	0	4,764	0.044	0.969	1.15×10^{-151}	1.04×10^{-149}
Thyroiditis (ICD10-10: E06)	*PHACTR2*	3	343	5	2,382	0	4,764	0.033	0.969	1.07×10^{-10}	1.92×10^{-10}
Males (Genetic sex)	*S100PBP*	48	125	687	2,382	11	4,764	0.022	0.078	0	0
Males (Genetic sex)	*ATP9A KCNB1*	55	144	687	2,382	31	4,764	0.016	0.078	7.29×10^{-251}	3.28×10^{-250}

Table 10: Frequency distribution of quantitative phenotypic features in cases associated with gene(s) compared to all cases and controls in the cohort. *p*-values were calculated to assess the association of each feature using two-sided Mann-Whitney U tests.

Phenotypic Feature	Gene(s)	Gene case count with feature data	Max, Mean & Min value of feature for cases with gene assoc.	All case count with feature data	Max, Mean & Min value of feature for all cases	Control count with gene assoc. and feature data	Max, Mean & Min value of feature for controls	*p*-value (compared to cases) before multiple testing	Adjusted *p*-value (compared to cases) after multiple testing
Phenylalanine levels in plasma	*SLC6A11 SULF2*	41	0.07 0.047 0.032	595	0.08 0.044 0.017	0	N/A	0.022	0.988

Table 11: Frequency distribution of quantitative phenotypic features in cases associated with communities compared to all cases and controls in the cohort. *p*-values were calculated to assess the association of each feature using two-sided Mann-Whitney U tests.

Phenotypic Feature	Community	Gene case count with feature data	Max, Mean & Min value of feature for cases with gene assoc.	All case count with feature data	Max, Mean & Min value of feature for all cases	Control count with gene assoc. and feature data	Max, Mean & Min value of feature for controls	*p*-value (compared to cases) before multiple testing	Adjusted *p*-value (compared to cases) after multiple testing
Lactate levels in plasma	3	61	1.81 3.41 6.81	595	1.41 3.69 9.64	0	N/A	0.028	0.992
Lactate levels in plasma	14	100	1.66 3.92 7.13	595	1.41 3.69 9.64	0	N/A	0.017	0.992

Genes Associated with SNPs using eQTL and Chromatin Interaction Data

Table 12: Gene assignments for SNPs using publicly available eQTL and chromatin-interaction data.

SNP	Genes assigned using eQTL/chromatin data	Gene assignment approach	Tissue(s) of Interest	Genes assigned from genome assembly
rs2904106	*SALL4*	eqtl	Brain_Cerebellum	*ATP9A*
rs3785477	*RP11-166P13.3*	eqtl	Whole_Blood, Brain_Hypothalamus, Brain_Frontal_Cortex_BA9, Brain_Putamen_basal_ganglia, Brain_Nucleus_accumbens_basal_ganglia	*AKAP1*
rs41306603	*FNDC5*	eqtl	Brain_Frontal_Cortex_BA9	*S100PBP*
rs6832769	*SRD5A3*	eqtl	Whole_Blood, Brain_Cerebellum	*CLOCK*
rs11839732	*LINC00346, TEX29*	promoter	Dorsolateral_Prefrontal_Cortex	-

Tissue Expression of Genes

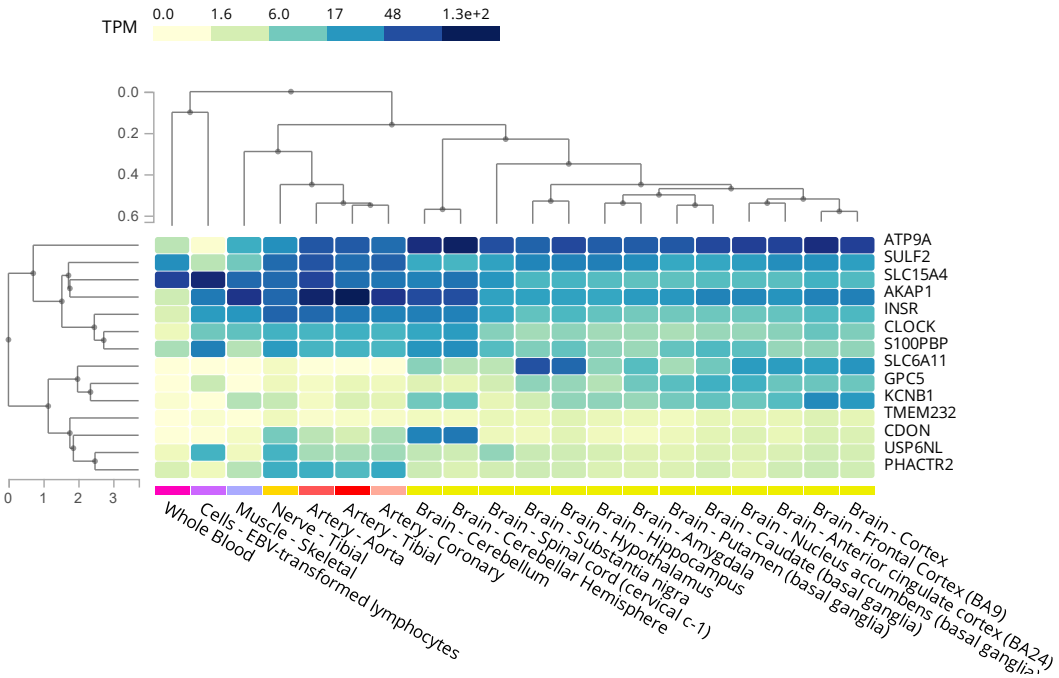


Figure 16: Clustered heatmap showing tissue expression profiles (GTEx) for 14 genes identified in the Pain Questionnaire study.

Comparison with UK Biobank CFS Verbal Interview Study

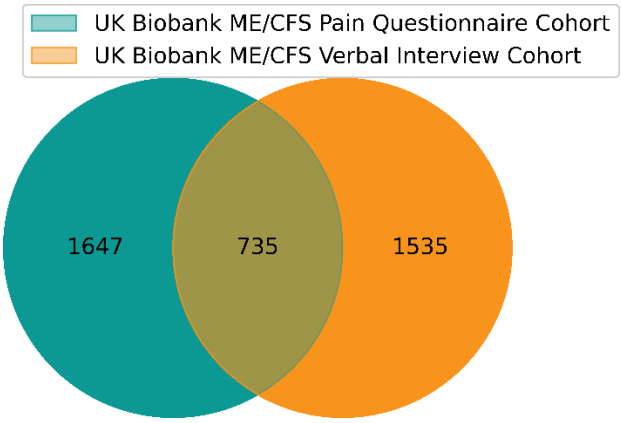


Figure 17: Case overlap between two UK Biobank ME/CFS cohorts (Pain Questionnaire and Verbal Interview)

Gene Annotation Data Sources

Table 13: Example public annotation sources for various types of study results

Annotation Type	Data Source(s)
Encoded Protein Structure and Function	Uniprot, InterPro, PDBe-KB
SNP-Disease Association	GWAS Catalog, PharmGKB, OpenTargets
Gene-Disease Association	PubMed, PubChem NCBI, OpenTargets
Gene Tissue/Cell Expression	Human Protein Atlas, GTEx, Promoter Capture HiC, Expression Atlas
Pathways, Interactions or MoA	Reactome, KEGG, Gene Ontology, PANTHER, WikiPathways, STRING, PubMed, IntAct
Safety / Toxicology	International Mouse Phenotyping Consortium, Tox21, eTox, HeCaToS, Open Targets
Drugs and chemical compounds	ChEMBL, DrugBank, PubChem NCBI, ProbeMiner, PDBe-KB

Table 14: Biological pathway enrichment results for genes associated with disease signatures identified in the ME/CFS Pain Questionnaire study using g:Profiler. Only significant results are reported from the enrichment analysis that used only non-electronic gene annotations, the g:SCS algorithm for multiple testing correction and considered genes with at least one annotation as background genes.

Disease Signature No.	Annotation source	Annotation name	Annotation ID	p_value	Genes in Disease Signatures (query genes)	Genes with Annotation
1	GO:BP	negative regulation of secretion by cell	GO:1903531	0.012606	AL035685.1, KCNB1, ATP9A	AL035685.1, KCNB1
1	GO:BP	negative regulation of secretion	GO:0051048	0.016488	AL035685.1, KCNB1, ATP9A	AL035685.1, KCNB1
1	GO:BP	regulation of exocytosis	GO:0017157	0.033035	AL035685.1, KCNB1, ATP9A	AL035685.1, KCNB1

6	WP	Ciliary landscape	WP:WP4352	0.025915	EXOC4, CEP290, SLC15A4	EXOC4, CEP290
6	GO:CC	specific granule	GO:0042581	0.036312	EXOC4, CEP290, SLC15A4	CEP290, SLC15A4
22	GO:BP	negative regulation of secretion by cell	GO:1903531	0.012606	C8orf34, AL035685.1, KCNB1, ATP9A	AL035685.1, KCNB1
22	GO:BP	negative regulation of secretion	GO:0051048	0.016488	C8orf34, AL035685.1, KCNB1, ATP9A	AL035685.1, KCNB1
22	GO:BP	regulation of exocytosis	GO:0017157	0.033035	C8orf34, AL035685.1, KCNB1, ATP9A	AL035685.1, KCNB1
23	GO:BP	negative regulation of secretion by cell	GO:1903531	0.012606	AP000920.1, AL035685.1, KCNB1, ATP9A	AP000920.1, AL035685.1
23	GO:BP	negative regulation of secretion	GO:0051048	0.016488	AP000920.1, AL035685.1, KCNB1, ATP9A	AP000920.1, AL035685.1
23	GO:BP	regulation of exocytosis	GO:0017157	0.033035	AP000920.1, AL035685.1, KCNB1, ATP9A	AP000920.1, AL035685.1
24	GO:BP	negative regulation of secretion by cell	GO:1903531	0.012606	AC009486.1, AL035685.1, KCNB1, ATP9A	AC009486.1, AL035685.1
24	GO:BP	negative regulation of secretion	GO:0051048	0.016488	AC009486.1, AL035685.1, KCNB1, ATP9A	AC009486.1, AL035685.1
24	GO:BP	regulation of exocytosis	GO:0017157	0.033035	AC009486.1, AL035685.1, KCNB1, ATP9A	AC009486.1, AL035685.1
30	REAC	Visual phototransduction	REAC:R-HSA-2187338	0.042912	NMS, GPC5, STRA6	GPC5, STRA6
31	REAC	Visual phototransduction	REAC:R-HSA-2187338	0.042912	NMS, SYNPR, SYNPR-AS1, GPC5, STRA6	GPC5, STRA6
47	GO:BP	negative regulation of secretion by cell	GO:1903531	0.012606	S100PBP, AC009486.1, AL035685.1, KCNB1, ATP9A	AC009486.1, AL035685.1
47	GO:BP	negative regulation of secretion	GO:0051048	0.016488	S100PBP, AC009486.1, AL035685.1, KCNB1, ATP9A	AC009486.1, AL035685.1
47	GO:BP	regulation of exocytosis	GO:0017157	0.033035	S100PBP, AC009486.1, AL035685.1, KCNB1, ATP9A	AC009486.1, AL035685.1
48	GO:BP	negative regulation of secretion by cell	GO:1903531	0.012606	AC009486.1, AP001978.1, AL035685.1,	AC009486.1, AP001978.1

					KCNB1, ATP9A	
48	GO:BP	negative regulation of secretion	GO:0051048	0.016488	AC009486.1, AP001978.1, AL035685.1, KCNB1, ATP9A	AC009486.1, AP001978.1
48	GO:BP	regulation of exocytosis	GO:0017157	0.033035	AC009486.1, AP001978.1, AL035685.1, KCNB1, ATP9A	AC009486.1, AP001978.1
49	GO:BP	regulation of Golgi organization	GO:1903358	0.003641	FAT4, USP6NL, CDON, CAMSAP3	USP6NL, CAMSAP3
59	KEGG	MAPK signaling pathway	KEGG:04010	0.036028	LINC01435, PTPRR, RN7SL363P, INSR	PTPRR, INSR
60	GO:BP	positive regulation of developmental growth	GO:0048639	0.038766	LRRC4C, AC090720.1, AP000446.1, SYT17, AC009097.2, INSR	AC090720.1, AP000446.1
63	REAC	GABA synthesis, release, reuptake and degradation	REAC:R-HSA-888590	0.000709	SLC6A11, AC027804.1, SYT1, AC090709.1, AC008738.6, SULF2	SLC6A11, AC027804.1
63	KEGG	Synaptic vesicle cycle	KEGG:04721	0.002323	SLC6A11, AC027804.1, SYT1, AC090709.1, AC008738.6, SULF2	SLC6A11, AC027804.1
63	REAC	Neurotransmitter release cycle	REAC:R-HSA-112310	0.005678	SLC6A11, AC027804.1, SYT1, AC090709.1, AC008738.6, SULF2	SLC6A11, AC027804.1

References

[1] Aoun Sebaiti M, Hainselin M, Gounden Y, Sirbu CA, Sekulic S, Lorusso L, Nacul L, Authier FJ. Systematic review and meta-analysis of cognitive impairment in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Sci Rep.* 2022 Feb 9;12(1):2157. doi: 10.1038/s41598-021-04764-w

[2] Nacul LC, Lacerda EM, Pheby D, Campion P, Molokhia M, Fayyaz S, Leite JC, Poland F, Howe A, Drachler ML. Prevalence of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) in three regions of England: a repeated cross-sectional study in primary care. *BMC Med.* 2011 Jul 28;9:91. doi: 10.1186/1741-7015-9-91

[3] Cortes Rivera M, Mastronardi C, Silva-Aldana CT, Arcos-Burgos M, Lidbury BA. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: A Comprehensive Review. *Diagnostics* (Basel). 2019 Aug 7;9(3):91. doi: 10.3390/diagnostics9030091

[4] Ruiz-Pablos M, Paiva B, Montero-Mateo R, Garcia N, Zabaleta A. Epstein-Barr Virus and the Origin of Myalgic Encephalomyelitis or Chronic Fatigue Syndrome. *Front Immunol.* 2021 Nov 15;12:656797. doi: 10.3389/fimmu.2021.656797

[5] Rasa S., Nora-Krukle Z., Henning N., Eliassen E., Shikova E., Harrer T., Scheibenbogen C., Murovska M., Prusty B.K., on behalf of the European Network on ME/CFS (EUROMEME) Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome. *J. Transl. Med.* 2018;10:268. doi: 10.1186/s12967-018-1644-y

[6] Hickie I., Davenport T., Wakefield D., Vollmer-Conna U., Cameron B., Vernon S.D., Reeves W.C., Lloyd A., for the Dubbo Infection Outcome Study Group Post-infective and chronic fatigue syndromes precipitated by viral and non-viral pathogens: Prospective cohort study. *BMJ.* 2006 doi: 10.1136/bmj.38933.585764.AE

[7] Katz B.Z., Shiraishi Y., Mears C.J., Binns H.S., Taylor R. Chronic fatigue syndrome after infectious mononucleosis in adolescents. *Pediatrics.* 2009;124:189–193. doi: 10.1542/peds.2008-1879]

[8] Chu L., Valencia I.J., Garvet D.W., Montoya J.G. Onset patterns and course of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Front. Pediatr.* 2019;7:12. doi: 10.3389/fped.2019.00012

[9] Balinas C, Eaton-Fitch N, Maksoud R, Staines D, Marshall-Gradisnik S. Impact of Life Stressors on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Symptoms: An Australian Longitudinal Study. *Int J Environ Res Public Health.* 2021 Oct 11;18(20):10614. doi: 10.3390/ijerph182010614

[10] Poenaru S, Abdallah SJ, Corrales-Medina V, Cowan J. COVID-19 and post-infectious myalgic encephalomyelitis/chronic fatigue syndrome: a narrative review. *Ther Adv Infect Dis.* 2021 Apr 20;8:20499361211009385. doi: 10.1177/20499361211009385

[11] Albright F, Light K, Light A, Bateman L, Cannon-Albright LA. Evidence for a heritable predisposition to Chronic Fatigue Syndrome. *BMC Neurol.* 2011 May 27;11:62. doi: 10.1186/1471-2377-11-62

[12] Dibble JJ, McGrath SJ, Ponting CP. Genetic risk factors of ME/CFS: a critical review. *Hum Mol Genet.* 2020 Sep 30;29(R1):R117-R124. doi: 10.1093/hmg/ddaa169

[13] Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet.* 2019 Aug;20(8):467-484. doi: 10.1038/s41576-019-0127-1

[14] Abell NS, DeGorter MK, Gloudemans MJ, Greenwald E, Smith KS, He Z, Montgomery SB. Multiple causal variants underlie genetic associations in humans. *Science*. 2022 Mar 18;375(6586):1247-1254. doi: 10.1126/science.abj5117

[15] Gardner, S. Combinatorial Analytics: An Essential Tool for the Delivery of Precision Medicine and Precision Agriculture 2021 *Artificial Intelligence in the Life Sciences*, 1, 100003 https://doi.org/10.1016/j.ailsci.2021.100003

[16] Koefoed, P., Andreassen, O.A., Bennike, B., Dam, H., Djurovic, S., Hansen, T., Jorgensen, M.B., Kessing, L.V., Melle, I., Møller, G.L., et al. (2011). Combinations of SNPs related to signal transduction in bipolar disorder. *PLoS One* 6, e23812. https://doi.org/10.1371/journal.pone.0023812

[17] Das, S., Pearson, M., Taylor, K., Bouchet, V., Møller, G.L., Hall, T.O., Strivens, M., Tzeng, K.T., and Gardner, S. (2021). Combinatorial analysis of phenotypic and clinical risk factors associated with hospitalized COVID-19 patients. *Front. Digit Health* 3, 660809. https://doi.org/10.3389/fdgth.2021.660809.

[18] Taylor, K., Das, S., Pearson, M., Kozubek, J., Strivens, M., and Gardner, S. (2019). Systematic drug repurposing to enable precision medicine: a case study in breast cancer. *Digital Med*. 5, 180. https://doi.org/10.4103/digm.digm_28_19.

[19] Taylor, K., Das, S., Pearson, M., Kozubek, J., Pawlowski, M., Jensen, C.E., Skowron, Z., Møller, G.L., Strivens, M., and Gardner, S. (2020). Analysis of genetic host response risk factors in severe COVID-19 patients. Preprint at medRxiv. https://doi.org/10.1101/2020.06.17.20134015.

[20] Shelton JF, Shastri AJ, Ye C, Weldon CH, Filshtein-Sonmez T, Coker D, Symons A, Esparza-Gordillo J; 23andMe COVID-19 Team, Aslibekyan S, Auton A. Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. Nat Genet. 2021 Jun;53(6):801-808. doi: 10.1038/s41588-021-00854-7

[21] COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature*. 2021 Dec;600(7889):472-477. doi: 10.1038/s41586-021-03767-x. Epub 2021 Jul 8.

[22] Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209. https://doi.org/10.1038/s41586-018-0579-z.

[23] Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple hypothesis testing. *J R Stat Soc B* 57:289–300

[24] Hajdarevic R, Lande A, Mehlsen J, Rydland A, Sosa DD, Strand EB, Mella O, Pociot F, Fluge Ø, Lie BA, Viken MK. Genetic association study in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) identifies several potential risk loci. *Brain Behav Immun.* 2022 May;102:362-369. doi: 10.1016/j.bbi.2022.03.010

[25] Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, Bhai J, et al. Ensembl 2021. *Nucleic Acids Res*. 2021 Jan 8;49(D1):D884-D891. doi: 10.1093/nar/gkaa942

[26]GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020 Sep 11;369(6509):1318-30.

[27] Jung I, Schmitt A, Diao Y, Lee AJ, Liu T, Yang D, Tan C, Eom J, Chan M, Chee S, Chiang Z. A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nature Genetics*. 2019 Oct;51(10):1442-9.

[28] Morgan P, Brown DG, Lennard S, Anderton MJ, Barrett JC, Eriksson U, Fidock M, Hamren B, Johnson A, March RE, Matcham J. Impact of a five-dimensional framework on R&D productivity at AstraZeneca. *Nature Reviews Drug Discovery*. 2018 Mar;17(3):167-81.

[29] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*. 2007 Sep 1;81(3):559-75.

[30] Qie H, Li S, Dou Y, Xu J, Xiong Y, Gao Z. Isolate sets partition benefits community detection of parallel Louvain method. *Sci Rep.* 2022 May 17;12(1):8248. doi: 10.1038/s41598-022-11987-y

[31] Gardinassi LG. A Cross-Study Biomarker Signature of Human Bronchial Epithelial Cells Infected with Respiratory Syncytial Virus. *Adv Virol.* 2016;2016:3605302. doi: 10.1155/2016/3605302. Epub 2016 May 4.

[32] Ansari IU, Longacre MJ, Paulusma CC, Stoker SW, Kendrick MA, MacDonald MJ. Characterization of P4 ATPase Phospholipid Translocases (Flippases) in Human and Rat Pancreatic Beta Cells: Their Gene Silencing Inhibits Insulin Secretion. *J Biol Chem*. 2015 Sep 18;290(38):23110-23. doi: 10.1074/jbc.M115.655027. Epub 2015 Aug 3.

[33] Fazia T, Marzanati D, Carotenuto AL, Beecham A, Hadjixenofontos A, McCauley JL, Saddi V, Piras M, Bernardinelli L, Gentilini D. Homozygosity Haplotype and Whole-Exome Sequencing Analysis to Identify Potentially Functional Rare Variants Involved in Multiple Sclerosis among Sardinian Families. *Curr Issues Mol Biol.* 2021 Oct 27;43(3):1778-1793. doi: 10.3390/cimb43030125.

[34] Li XN, Herrington J, Petrov A, Ge L, Eiermann G, Xiong Y, Jensen MV, Hohmeier HE, Newgard CB, Garcia ML, Wagner M, Zhang BB, Thornberry NA, Howard AD, Kaczorowski GJ, Zhou YP. The role of voltage-gated potassium channels Kv2.1 and Kv2.2 in the regulation of insulin and somatostatin release from pancreatic islets. *J Pharmacol Exp Ther.* 2013 Feb;344(2):407-16. doi: 10.1124/jpet.112.199083. Epub 2012 Nov 16.

[35] Huang W, Ramsey KM, Marcheva B, Bass J. Circadian rhythms, sleep, and metabolism. *J Clin Invest.* 2011 Jun;121(6):2133-41. doi: 10.1172/JCI46043. Epub 2011 Jun 1.

[36] de Goede P, Wefers J, Brombacher EC, Schrauwen P, Kalsbeek A. Circadian rhythms in mitochondrial respiration. *J Mol Endocrinol.* 2018 Apr;60(3):R115-R130. doi: 10.1530/JME-17-0196. Epub 2018 Jan 29.

[37] Stenvers DJ, Scheer FAJL, Schrauwen P, la Fleur SE, Kalsbeek A. Circadian clocks and insulin resistance. *Nat Rev Endocrinol.* 2019 Feb;15(2):75-89. doi: 10.1038/s41574-018-0122-1.

[38] Kobayashi T, Shimabukuro-Demoto S, Yoshida-Sugitani R, Furuyama-Tanaka K, Karyu H, Sugiura Y, Shimizu Y, Hosaka T, Goto M, Kato N, Okamura T, Suematsu M, Yokoyama S, Toyama-Sorimachi N. The histidine transporter SLC15A4 coordinates mTOR-dependent inflammatory responses and pathogenic antibody production. *Immunity*. 2014 Sep 18;41(3):375-388. doi: 10.1016/j.immuni.2014.08.011.

[39] Kobayashi T, Nguyen-Tien D, Ohshima D, Karyu H, Shimabukuro-Demoto S, Yoshida-Sugitani R, Toyama-Sorimachi N. Human SLC15A4 is crucial for TLR-mediated type I interferon production and mitochondrial integrity. *Int Immunol.* 2021 Jun 18;33(7):399-406. doi: 10.1093/intimm/dxab006.

[40] Kobayashi T, Nguyen-Tien D, Sorimachi Y, Sugiura Y, Suzuki T, Karyu H, Shimabukuro-Demoto S, Uemura T, Okamura T, Taguchi T, Ueki K, Kato N, Goda N, Dohmae N, Takubo K, Suematsu M, Toyama-Sorimachi N. SLC15A4 mediates M1-prone metabolic shifts in macrophages and guards immune cells from metabolic stress. *Proc Natl Acad Sci U S A.* 2021 Aug 17;118(33):e2100295118. doi: 10.1073/pnas.2100295118.

[41] Souren NY, Gerdes LA, Lutsik P, Gasparoni G, Beltrán E, Salhab A, Kümpfel T, Weichenhan D, Plass C, Hohlfeld R, Walter J. DNA methylation signatures of monozygotic twins clinically discordant for multiple sclerosis. *Nat Commun*. 2019 May 7;10(1):2094. doi: 10.1038/s41467-019-09984-3.

[42] Chorąży M, Wawrusiewicz-Kurylonek N, Posmyk R, Zajkowska A, Kapica-Topczewska K, Krętowski AJ, Kochanowicz J, Kułakowska A. Analysis of chosen SNVs in GPC5, CD58 and IRF8 genes in multiple sclerosis patients. *Adv Med Sci.* 2019 Sep;64(2):230-234. doi: 10.1016/j.advms.2018.12.004. Epub 2019 Feb 26.

[43] Mowry EM, Carey RF, Blasco MR, Pelletier J, Duquette P, Villoslada P, Malikova I, Roger E, Kinkel RP, McDonald J, Bacchetti P, Waubant E. Multiple sclerosis susceptibility genes: associations with relapse severity and recovery. *PLoS One*. 2013 Oct 9;8(10):e75416. doi: 10.1371/journal.pone.0075416. PMID: 24130709;

[44] Goldstein BA, Hubbard AE, Cutler A, Barcellos LF. An application of Random Forests to a genome-wide association dataset: methodological considerations & new findings. *BMC Genet.* 2010 Jun 14;11:49. doi: 10.1186/1471-2156-11-49.

[45] Schiattarella GG, Cattaneo F, Carrizzo A, Paolillo R, Boccella N, Ambrosio M, Damato A, Pironti G, Franzone A, Russo G, Magliulo F, Pirozzi M, Storto M, Madonna M, Gargiulo G, Trimarco V, Rinaldi L, De Lucia M, Garbi C, Feliciello A, Esposito G, Vecchione C, Perrino C. *Akap1* Regulates Vascular Function and Endothelial Cells Behavior. *Hypertension*. 2018 Mar;71(3):507-517. doi: 10.1161/HYPERTENSIONAHA.117.10185. Epub 2018 Jan 15.

[46] Narala VR, Fukumoto J, Hernández-Cuervo H, Patil SS, Krishnamurthy S, Breitzig M, Galam L, Soundararajan R, Lockey RF, Kolliputi N. Akap1 genetic deletion increases the severity of hyperoxia-induced acute lung injury in mice. *Am J Physiol Lung Cell Mol Physiol.* 2018 May 1;314(5):L860-L870. doi: 10.1152/ajplung.00365.2017. Epub 2018 Feb 1.

[47] Zenner HL, Yoshimura S, Barr FA, Crump CM. Analysis of Rab GTPase-activating proteins indicates that Rab1a/b and Rab43 are important for herpes simplex virus 1 secondary envelopment. *J Virol.* 2011 Aug;85(16):8012-21. doi: 10.1128/JVI.00500-11. Epub 2011 Jun 15.

[48] Bae JH, Hong M, Jeong HJ, Kim H, Lee SJ, Ryu D, Bae GU, Cho SC, Lee YS, Krauss RS, Kang JS. Satellite cell-specific ablation of Cdon impairs integrin activation, FGF signalling, and muscle regeneration. *J Cachexia Sarcopenia Muscle.* 2020 Aug;11(4):1089-1103. doi: 10.1002/jcsm.12563. Epub 2020 Feb 27. Erratum in: *J Cachexia Sarcopenia Muscle*. 2020 Oct;11(5):1381.

[49] Wang LC, Almazan G. Cdon, a cell surface protein, mediates oligodendrocyte differentiation and myelination. *Glia*. 2016 Jun;64(6):1021-33. doi: 10.1002/glia.22980. Epub 2016 Mar 14.

[50] Shukla SK, Rose W, Schrodi SJ. Complex host genetic susceptibility to Staphylococcus aureus infections. *Trends Microbiol.* 2015 Sep;23(9):529-36. doi: 10.1016/j.tim.2015.05.008. Epub 2015 Jun 22.

[51] Højlund K. Metabolism and insulin signaling in common metabolic disorders and inherited insulin resistance. *Dan Med J.* 2014 Jul;61(7):B4890.

[52] Ye J, Wen Y, Chu X, Li P, Cheng B, Cheng S, Liu L, Zhang L, Ma M, Qi X, Liang C, Kafle OP, Jia Y, Wu C, Wang S, Wang X, Ning Y, Zhang F. Association between herpes simplex virus 1 exposure and the risk of depression in UK Biobank. *Clin Transl Med.* 2020 Jun;10(2):e108. doi: 10.1002/ctm2.108. Epub 2020 Jun 20.

[53] Hassing HC, Surendran RP, Derudas B, Verrijken A, Francque SM, Mooij HL, Bernelot Moens SJ, Hart LM, Nijpels G, Dekker JM, Williams KJ, Stroes ES, Van Gaal LF, Staels B, Nieuwdorp M, Dallinga-Thie GM. SULF2 strongly prediposes to fasting and postprandial triglycerides in patients with obesity and type 2 diabetes mellitus. *Obesity (Silver Spring).* 2014 May;22(5):1309-16. doi: 10.1002/oby.20682. Epub 2014 Jan 9.

[54] Narita M, Niikura K, Nanjo-Niikura K, Narita M, Furuya M, Yamashita A, Saeki M, Matsushima Y, Imai S, Shimizu T, Asato M, Kuzumaki N, Okutsu D, Miyoshi K, Suzuki M, Tsukiyama Y, Konno M, Yomiya K, Matoba M, Suzuki T. Sleep disturbances in a neuropathic pain-like condition in the mouse are associated with altered GABAergic transmission in the cingulate cortex. *Pain*. 2011 Jun;152(6):1358-1372. doi: 10.1016/j.pain.2011.02.016. Epub 2011 Mar 10.

[55] Yamashita A, Hamada A, Suhara Y, Kawabe R, Yanase M, Kuzumaki N, Narita M, Matsui R, Okano H, Narita M. Astrocytic activation in the anterior cingulate cortex is critical for sleep disorder under neuropathic pain. *Synapse*. 2014 Jun;68(6):235-47. doi: 10.1002/syn.21733. Epub 2014 Feb 24..

[56] Kammel LG, Wei W, Jami SA, Voskuhl RR, O'Dell TJ. Enhanced GABAergic Tonic Inhibition Reduces Intrinsic Excitability of Hippocampal CA1 Pyramidal Cells in Experimental Autoimmune Encephalomyelitis. *Neuroscience*. 2018 Dec 15;395:89-100. doi: 10.1016/j.neuroscience.2018.11.003. Epub 2018 Nov 14. PMID: 30447391;

[57] Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J. g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic acids research*. 2019 Jul 2;47(W1):W191-8.

[58] Hickie I, Davenport T, Wakefield D, Vollmer-Conna U, Cameron B, Vernon SD, Reeves WC, Lloyd A; Dubbo Infection Outcomes Study Group. Post-infective and chronic fatigue syndromes precipitated by viral and non-viral pathogens: prospective cohort study. *BMJ*. 2006 Sep 16;333(7568):575. doi: 10.1136/bmj.38933.585764.AE

[59] Raijmakers RPH, Roerink ME, Jansen AFM, Keijmel SP, Gacesa R, Li Y, Joosten LAB, van der Meer JWM, Netea MG, Bleeker-Rovers CP, Xu CJ. Multi-omics examination of Q fever fatigue syndrome identifies similarities with chronic fatigue syndrome. *J Transl Med.* 2020 Nov 26;18(1):448. doi: 10.1186/s12967-020-02585-5

[60] Wong TL, Weitzer DJ. Long COVID and Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS)-A Systemic Review and Comparison of Clinical Presentation and Symptomatology. *Medicina (Kaunas).* 2021 Apr 26;57(5):418. doi: 10.3390/medicina57050418

[61] Morris G, Berk M, Galecki P, Maes M. The emerging role of autoimmunity in myalgic encephalomyelitis/chronic fatigue syndrome (ME/cfs). *Mol Neurobiol.* 2014 Apr;49(2):741-56. doi: 10.1007/s12035-013-8553-0

[62] Deumer US, Varesi A, Floris V, Savioli G, Mantovani E, López-Carrasco P, Rosati GM, Prasad S, Ricevuti G. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS): An Overview. *J Clin Med.* 2021 Oct 19;10(20):4786. doi: 10.3390/jcm10204786

[63] Rasa S, Nora-Krukle Z, Henning N, Eliassen E, Shikova E, Harrer T, Scheibenbogen C, Murovska M, Prusty BK; European Network on ME/CFS (EUROMENE). Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *J Transl Med.* 2018 Oct 1;16(1):268. doi: 10.1186/s12967-018-1644-y

[64] Morris G, Maes M. Mitochondrial dysfunctions in myalgic encephalomyelitis/chronic fatigue syndrome explained by activated immuno-inflammatory, oxidative and nitrosative stress pathways. *Metab Brain Dis.* 2014 Mar;29(1):19-36. doi: 10.1007/s11011-013-9435-x

[65] Barrera MJ, Aguilera S, Castro I, Carvajal P, Jara D, Molina C, González S, González MJ. Dysfunctional mitochondria as critical players in the inflammation of autoimmune diseases: Potential role in Sjögren's syndrome. *Autoimmun Rev.* 2021 Aug;20(8):102867. doi: 10.1016/j.autrev.2021.102867

[66] Clayton SA, MacDonald L, Kurowska-Stolarska M, Clark AR. Mitochondria as Key Players in the Pathogenesis and Treatment of Rheumatoid Arthritis. *Front Immunol*. 2021 Apr 29;12:673916. doi: 10.3389/fimmu.2021.673916

[67] Wang C, Ahlford A, Järvinen TM, Nordmark G, Eloranta ML, Gunnarsson I, Svenungsson E, Padyukov L, Sturfelt G, Jönsen A, Bengtsson AA, Truedsson L, Eriksson C, Rantapää-Dahlqvist S, Sjöwall C, Julkunen H, Criswell LA, Graham RR, Behrens TW, Kere J, Rönnblom L, Syvänen AC, Sandling JK. Genes identified in Asian SLE GWASs are also associated with SLE in Caucasian populations. *Eur J Hum Genet.* 2013 Sep;21(9):994-9. doi: 10.1038/ejhg.2012.277

[68] Kobayashi T, Nguyen-Tien D, Sorimachi Y, Sugiura Y, Suzuki T, Karyu H, Shimabukuro-Demoto S, Uemura T, Okamura T, Taguchi T, Ueki K, Kato N, Goda N, Dohmae N, Takubo K, Suematsu M, Toyama-Sorimachi N. SLC15A4 mediates M1-prone metabolic shifts in macrophages and guards immune cells from metabolic stress. *Proc Natl Acad Sci* U S A. 2021 Aug 17;118(33):e2100295118. doi: 10.1073/pnas.2100295118

[69] Kobayashi T, Nguyen-Tien D, Ohshima D, Karyu H, Shimabukuro-Demoto S, Yoshida-Sugitani R, Toyama-Sorimachi N. Human SLC15A4 is crucial for TLR-mediated type I interferon production and mitochondrial integrity. *Int Immunol.* 2021 Jun 18;33(7):399-406. doi: 10.1093/intimm/dxab006

[70] Shin JG, Kim HJ, Park BL, Bae JS, Kim LH, Cheong HS, Shin HD. Putative association of GPC5 polymorphism with the risk of inflammatory demyelinating diseases. *J Neurol Sci.* 2013 Dec 15;335(1-2):82-8. doi: 10.1016/j.jns.2013.08.031

[71] Chorąży M, Wawrusiewicz-Kurylonek N, Posmyk R, Zajkowska A, Kapica-Topczewska K, Krętowski AJ, Kochanowicz J, Kułakowska A. Analysis of chosen SNVs in GPC5, CD58 and IRF8 genes in multiple sclerosis patients. *Adv Med Sci.* 2019 Sep;64(2):230-234. doi: 10.1016/j.advms.2018.12.004

[72] Johnson BA, Wang J, Taylor EM, Caillier SJ, Herbert J, Khan OA, Cross AH, De Jager PL, Gourraud PA, Cree BC, Hauser SL, Oksenberg JR. Multiple sclerosis susceptibility alleles in African Americans. *Genes Immun.* 2010 Jun;11(4):343-50. doi: 10.1038/gene.2009.81

[73] Jain V, Arunkumar A, Kingdon C, Lacerda E, Nacul L. Prevalence of and risk factors for severe cognitive and sleep symptoms in ME/CFS and MS. *BMC Neurol.* 2017 Jun 20;17(1):117. doi: 10.1186/s12883-017-0896-0

[74] Tomas C, Brown A, Strassheim V, Elson JL, Newton J, Manning P. Cellular bioenergetics is impaired in patients with chronic fatigue syndrome. *PLoS One*. 2017 Oct 24;12(10):e0186802. doi: 10.1371/journal.pone.0186802. Erratum in: *PLoS One*. 2018 Feb 8;13(2):e0192817

[75] Liu Y, Merrill RA, Strack S. A-Kinase Anchoring Protein 1: Emerging Roles in Regulating Mitochondrial Form and Function in Health and Disease. *Cells*. 2020 Jan 26;9(2):298. doi: 10.3390/cells9020298

[76] Schiattarella GG, Cattaneo F, Carrizzo A, Paolillo R, Boccella N, Ambrosio M, Damato A, Pironti G, Franzone A, Russo G, Magliulo F, Pirozzi M, Storto M, Madonna M, Gargiulo G, Trimarco V, Rinaldi L, De Lucia M, Garbi C, Feliciello A, Esposito G, Vecchione C, Perrino C. Akap1 Regulates Vascular Function and Endothelial Cells Behavior. *Hypertension.* 2018 Mar;71(3):507-517. doi: 10.1161/HYPERTENSIONAHA.117.10185

[77] Yoshinaka T, Kosako, H., Yoshizumi, T., Furukawa, R., Hirano, Y., Kuge, O., Tamada, T., and Koshiba. T., Structural Basis of Mitochondrial Scaffolds by Prohibitin Complexes: Insight into a Role of the Coiled-Coil Region, *iScience* 19, 1065–1078 Sep 27, 2019 https://doi.org/10.1016/j.isci.2019.08.056

[78] Xu X, Xu L, Zhang P, Ouyang K, Xiao Y, Xiong J, Wang D, Liang Y, Duan L. Effects of ATP9A on Extracellular Vesicle Release and Exosomal Lipid Composition. *Oxid Med Cell Longev*. 2020 Oct 29;2020:8865499. doi: 10.1155/2020/8865499

[79] Nikolova-Karakashian MN, Reid MB. Sphingolipid metabolism, oxidant signaling, and contractile function of skeletal muscle. *Antioxid Redox Signal*. 2011 Nov 1;15(9):2501-17. doi: 10.1089/ars.2011.3940

[80] Che X, Brydges CR, Yu Y, Price A, Joshi S, Roy A, Lee B, Barupal DK, Cheng A, Palmer DM, Levine S, Peterson DL, Vernon SD, Bateman L, Hornig M, Montoya JG, Komaroff AL, Fiehn O, Lipkin WI. Evidence for Peroxisomal Dysfunction and Dysregulation of the CDP-Choline Pathway in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. medRxiv [Preprint]. 2022 Jan 11:2021.06.14.21258895. doi: 10.1101/2021.06.14.21258895

[81] Nagy-Szakal D, Barupal DK, Lee B, Che X, Williams BL, Kahn EJR, Ukaigwe JE, Bateman L, Klimas NG, Komaroff AL, Levine S, Montoya JG, Peterson DL, Levin B, Hornig M, Fiehn O, Lipkin WI. Insights into myalgic encephalomyelitis/chronic fatigue syndrome phenotypes through comprehensive metabolomics. *Sci Rep.* 2018 Jul 3;8(1):10056. doi: 10.1038/s41598-018-28477-9

[82] Ansari IU, Longacre MJ, Paulusma CC, Stoker SW, Kendrick MA, MacDonald MJ. Characterization of P4 ATPase Phospholipid Translocases (Flippases) in Human and Rat Pancreatic Beta Cells: Their Gene Silencing Inhibits Insulin Secretion. *J Biol Chem*. 2015 Sep 18;290(38):23110-23. doi: 10.1074/jbc.M115.655027

[83] Fazia T, Marzanati D, Carotenuto AL, Beecham A, Hadjixenofontos A, McCauley JL, Saddi V, Piras M, Bernardinelli L, Gentilini D. Homozygosity Haplotype and Whole-Exome Sequencing Analysis to Identify Potentially Functional Rare Variants Involved in Multiple Sclerosis among Sardinian Families. *Curr Issues Mol Biol*. 2021 Oct 27;43(3):1778-1793. doi: 10.3390/cimb43030125

[84] Allain TJ, Bearn JA, Coskeran P, Jones J, Checkley A, Butler J, Wessely S, Miell JP. Changes in growth hormone, insulin, insulinlike growth factors (IGFs), and IGF-binding protein-1 in chronic fatigue syndrome. *Biol Psychiatry*. 1997 Mar 1;41(5):567-73. doi: 10.1016/s0006-3223(96)00074-1

[85] Wirth KJ, Scheibenbogen C. Pathophysiology of skeletal muscle disturbances in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *J Transl Med*. 2021 Apr 21;19(1):162. doi: 10.1186/s12967-021-02833-2

[86] Choi CS, Kim YB, Lee FN, Zabolotny JM, Kahn BB, Youn JH. Lactate induces insulin resistance in skeletal muscle by suppressing glycolysis and impairing insulin signaling. *Am J Physiol Endocrinol Metab*. 2002 Aug;283(2):E233-40. doi: 10.1152/ajpendo.00557.2001

[87] Weyrauch LA, McMillin SL, Witczak CA. Insulin Resistance Does Not Impair Mechanical Overload-Stimulated Glucose Uptake, but Does Alter the Metabolic Fate of Glucose in Mouse Muscle. *Int J Mol Sci*. 2020 Jul 1;21(13):4715. doi: 10.3390/ijms21134715

[88] Burkart AM, Tan K, Warren L, Iovino S, Hughes KJ, Kahn CR, Patti ME. Insulin Resistance in Human iPS Cells Reduces Mitochondrial Size and Function. *Sci Rep.* 2016 Mar 7;6:22788. doi: 10.1038/srep22788

[89] Hernandez-Rabaza V, Cabrera-Pastor A, Taoro-Gonzalez L, Gonzalez-Usano A, Agusti A, Balzano T, Llansola M, Felipo V. Neuroinflammation increases GABAergic tone and impairs cognitive and motor function in hyperammonemia by increasing GAT-3 membrane expression. Reversal by sulforaphane by promoting M2 polarization of microglia. *J Neuroinflammation.* 2016 Apr 18;13(1):83. doi: 10.1186/s12974-016-0549-z

[90] Narita M, Niikura K, Nanjo-Niikura K, Narita M, Furuya M, Yamashita A, Saeki M, Matsushima Y, Imai S, Shimizu T, et al. Sleep disturbances in a neuropathic pain-like condition in the mouse are associated with altered GABAergic transmission in the cingulate cortex. *Pain*. 2011 Jun;152(6):1358-1372. doi: 10.1016/j.pain.2011.02.016

[91] Yamashita A, Hamada A, Suhara Y, Kawabe R, Yanase M, Kuzumaki N, Narita M, Matsui R, Okano H, Narita M. Astrocytic activation in the anterior cingulate cortex is critical for sleep disorder under neuropathic pain. *Synapse*. 2014 Jun;68(6):235-47. doi: 10.1002/syn.21733

[92] Zink M, Vollmayr B, Gebicke-Haerter PJ, Henn FA. Reduced expression of GABA transporter GAT3 in helpless rats, an animal model of depression. *Neurochem Res.* 2009 Sep;34(9):1584-93. doi: 10.1007/s11064-009-9947-2

[93] Albrecht A, Ivens S, Papageorgiou IE, Çalışkan G, Saiepour N, Brück W, Richter-Levin G, Heinemann U, Stork O. Shifts in excitatory/inhibitory balance by juvenile stress: A role for neuron-astrocyte interaction in the dentate gyrus. *Glia*. 2016 Jun;64(6):911-22. doi: 10.1002/glia.22970

[94] Kalus I, Rohn S, Puvirajesinghe TM, Guimond SE, Eyckerman-Kölln PJ, Ten Dam G, van Kuppevelt TH, Turnbull JE, Dierks T. Sulf1 and Sulf2 Differentially Modulate Heparan Sulfate Proteoglycan Sulfation during Postnatal Cerebellum Development: Evidence for Neuroprotective and Neurite Outgrowth Promoting Functions. *PLoS One*. 2015 Oct 8;10(10):e0139853. doi: 10.1371/journal.pone.0139853

[95] Kalus I, Salmen B, Viebahn C, von Figura K, Schmitz D, D'Hooge R, Dierks T. Differential involvement of the extracellular 6-O-endosulfatases Sulf1 and Sulf2 in brain development and neuronal and behavioural plasticity. *J Cell Mol Med.* 2009 Nov-Dec;13(11-12):4505-21. doi: 10.1111/j.1582-4934.2008.00558.x

[96] Joy MT, Vrbova G, Dhoot GK, Anderson PN. Sulf1 and Sulf2 expression in the nervous system and its role in limiting neurite outgrowth in vitro. *Exp Neurol.* 2015 Jan;263:150-60. doi: 10.1016/j.expneurol.2014.10.011

[97] Ye J, Wen Y, Chu X, Li P, Cheng B, Cheng S, Liu L, Zhang L, Ma M, Qi X, Liang C, Kafle OP, Jia Y, Wu C, Wang S, Wang X, Ning Y, Zhang F. Association between herpes simplex virus 1 exposure and the risk of depression in UK Biobank. *Clin Transl Med.* 2020 Jun;10(2):e108. doi: 10.1002/ctm2.108

[98] Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature Genetics.* 2018 Sep;50(9):1335-1341. DOI: 10.1038/s41588-018-0184-y. PMID: 30104761; PMCID: PMC6119127.

[99] Ye J, Wen Y, Chu X, Li P, Cheng B, Cheng S, Liu L, Zhang L, Ma M, Qi X, Liang C, Kafle OP, Jia Y, Wu C, Wang S, Wang X, Ning Y, Zhang F. Association between herpes simplex virus 1 exposure and the risk of depression in UK Biobank. *Clin Transl Med.* 2020 Jun;10(2):e108. doi: 10.1002/ctm2.108

[100] Bae JH, Hong M, Jeong HJ, Kim H, Lee SJ, Ryu D, Bae GU, Cho SC, Lee YS, Krauss RS, Kang JS. Satellite cell-specific ablation of Cdon impairs integrin activation, FGF signalling, and muscle regeneration. *J Cachexia Sarcopenia Muscle.* 2020 Aug;11(4):1089-1103. doi: 10.1002/jcsm.12563

[101] Shukla SK, Rose W, Schrodi SJ. Complex host genetic susceptibility to Staphylococcus aureus infections. *Trends Microbiol*. 2015 Sep;23(9):529-36. doi: 10.1016/j.tim.2015.05.008

[102] Verwey M, Grant A, Meti N, Adye-White L, Torres-Berrío A, Rioux V, Lévesque M, Charron F, Flores C. Mesocortical Dopamine Phenotypes in Mice Lacking the Sonic Hedgehog Receptor Cdon. *eNeuro*. 2016 Jul 13;3(3):ENEURO.0009-16.2016. doi: 10.1523/ENEURO.0009-16.2016

[103] Tomas C, Brown A, Strassheim V, Elson JL, Newton J, Manning P. Cellular bioenergetics is impaired in patients with chronic fatigue syndrome. *PLoS One*. 2017 Oct 24;12(10):e0186802. doi: 10.1371/journal.pone.0186802. Erratum in: PLoS One. 2018 Feb 8;13(2):e0192817

[104] de Goede P, Wefers J, Brombacher EC, Schrauwen P, Kalsbeek A. Circadian rhythms in mitochondrial respiration. *J Mol Endocrinol.* 2018 Apr;60(3):R115-R130. doi: 10.1530/JME-17-0196

[105] Schmitt K, Grimm A, Dallmann R, Oettinghaus B, Restelli LM, Witzig M, Ishihara N, Mihara K, Ripperger JA, Albrecht U, Frank S, Brown SA, Eckert A. Circadian Control of DRP1 Activity Regulates Mitochondrial Dynamics and Bioenergetics. *Cell Metab.* 2018 Mar 6;27(3):657-666.e5. doi: 10.1016/j.cmet.2018.01.011

[106] Oosterman JE, Wopereis S, Kalsbeek A. The Circadian Clock, Shift Work, and Tissue-Specific Insulin Resistance. *Endocrinology*. 2020 Dec 1;161(12):bqaa180. doi: 10.1210/endocr/bqaa180

[107] Sweetman E, Ryan M, Edgar C, MacKay A, Vallings R, Tate W. Changes in the transcriptome of circulating immune cells of a New Zealand cohort with myalgic encephalomyelitis/chronic fatigue syndrome. *Int J Immunopathol Pharmacol.* 2019 Jan-Dec;33:2058738418820402. doi: 10.1177/2058738418820402

[108] Natelson BH. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome and Fibromyalgia: Definitions, Similarities, and Differences. *Clin Ther.* 2019 Apr;41(4):612-618. doi: 10.1016/j.clinthera.2018.12.016

[109] Gatto N, Dos Santos Souza C, Shaw AC, Bell SM, Myszczynska MA, Powers S, Meyer K, Castelli LM, Karyka E, Mortiboys H, Azzouz M, Hautbergue GM, Márkus NM, Shaw PJ, Ferraiuolo L. Directly converted astrocytes retain the ageing features of the donor fibroblasts and elucidate the astrocytic contribution to human CNS health and disease. *Aging Cell.* 2021 Jan;20(1):e13281. doi: 10.1111/acel.13281

[110] Ochoa D, Hercules A, Carmona M, Suveges D, Gonzalez-Uriarte A, Malangone C, Miranda A, Fumis L, Carvalho-Silva D, Spitzer M, et al. Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res.* 2021 Jan 8;49(D1):D1302-D1310. doi: 10.1093/nar/gkaa1027

[111111] Jason LA, Ohanian D, Brown A, Sunnquist M, McManimen S, Klebek L, Fox P, Sorenson M. Differentiating multiple sclerosis from myalgic encephalomyelitis and chronic fatigue syndrome. *Insights in biomedicine*. 2017;2(2).

[112] Sukocheva OA, Maksoud R, Beeraka NM, Madhunapantula SV, Sinelnikov M, Nikolenko VN, Neganova ME, Klochkov SG, Kamal MA, Staines DR, Marshall-Gradisnik S. Analysis of post COVID-19 condition and its overlap with myalgic encephalomyelitis/chronic fatigue syndrome. *Journal of Advanced Research*. 2021 Nov 26.

[113] Chu L, Valencia IJ, Garvert DW, Montoya JG. Onset Patterns and Course of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Front Pediatr.* 2019 Feb 5;7:12. doi: 10.3389/fped.2019.00012

[114] Bakken IJ, Tveito K, Gunnes N, Ghaderi S, Stoltenberg C, Trogstad L, Håberg SE, Magnus P. Two age peaks in the incidence of chronic fatigue syndrome/myalgic encephalomyelitis: a population-based registry study from Norway 2008-2012. *BMC Med.* 2014 Oct 1;12:167. doi: 10.1186/s12916-014-0167-5

[115] Hewitt J, Walters M, Padmanabhan S, Dawson J. Cohort profile of the UK Biobank: diagnosis and characteristics of cerebrovascular disease. *BMJ Open*. 2016 Mar 22;6(3):e009161. doi: 10.1136/bmjopen-2015-009161

[116] Taylor AE, Jones HJ, Sallis H, Euesden J, Stergiakouli E, Davies NM, Zammit S, Lawlor DA, Munafò MR, Davey Smith G, Tilling K. Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol.* 2018 Aug 1;47(4):1207-1216. doi: 10.1093/ije/dyy060

[117] Jain V, Arunkumar A, Kingdon C, Lacerda E, Nacul L. Prevalence of and risk factors for severe cognitive and sleep symptoms in ME/CFS and MS. *BMC Neurol.* 2017 Jun 20;17(1):117. doi: 10.1186/s12883-017-0896-0

[118] Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, Elledge SJ, Niebuhr DW, Scher AI, Munger KL, Ascherio A. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science*. 2022 Jan 21;375(6578):296-301. doi: 10.1126/science.abj8222

[119] Komaroff AL, Bateman L. Will COVID-19 Lead to Myalgic Encephalomyelitis/Chronic Fatigue Syndrome? *Front Med* (Lausanne). 2021 Jan 18;7:606824. doi: 10.3389/fmed.2020.606824

[120] Kusama Y, Fukui S, Maruyama M, Kamimura K, Maihara T. Myalgic encephalomyelitis/chronic fatigue syndrome post coronavirus disease 2019. *Pediatr Int.* 2022 Jan;64(1):e14976. doi: 10.1111/ped.14976

[121] Poenaru S, Abdallah SJ, Corrales-Medina V, Cowan J. COVID-19 and post-infectious myalgic encephalomyelitis/chronic fatigue syndrome: a narrative review. *Ther Adv Infect Dis.* 2021 Apr 20;8:20499361211009385. doi: 10.1177/20499361211009385

[122] COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature* Jul 2021 600, 472–477.

[123] https://www.decodeme.org.uk/

[124] Das S, Taylor K, Beaulah S, Gardner S. Systematic indication extension for drugs using patient stratification insights generated by combinatorial analytics. *Patterns (N Y).* 2022 Jun 10;3(6):100496. doi: 10.1016/j.patter.2022.100496.

[125] Jin Y, Schaffer AA, Feolo M, Holmes JB, Kattman BL. GRAF-pop: a fast distance-based method to infer subject ancestry from multiple genotype datasets without principal components analysis. G3: Genes, Genomes, *Genetics*. 2019 Aug 1;9(8):2447-61.

[126] Jin Y, Schäffer AA, Sherry ST, and Feolo M (2017). Quickly identifying identical and closely related subjects in large databases using genotype data. *PLoS One.* 12(6):e0179106.