RESEARCH

Genetic risk factors for ME/CFS identified using combinatorial analysis

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

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 to 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 1000 cycles of fully random permutation. Results from this study were supported by a series of replication and cohort comparison experiments, including use of disjoint Verbal Interview CFS, post-viral fatigue syndrome and fibromyalgia cohorts also derived from UK Biobank, and compared results for overlap and 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 and 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 identified 3 of the critical SNPs replicated in the post-viral fatigue syndrome cohort and 2 SNPs replicated in the fibromyalgia cohort. We also 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 [4]. ME/CFS development has been associated with prior viral infection such as with Epstein-Barr Virus (EBV) [5] and other pathogens [6–9], however there is also evidence

[†]Sayoni Das, Krystyna Taylor and Steve Gardner have contributed equally

*Correspondence: steve@precisionlife.com

PrecisionLife Ltd, Long Hanborough, Oxford, UK

Das *et al. Journal of Translational Medicine*   (2022) 20:598

Page 2 of 20

that stress and non-viral infection may also contribute to triggering ME/CFS onset [10].

The multi-factorial spectrum of ME/CFS triggers and symptoms [11] 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 [12]. While there is a demonstrable heritable component to the disease [13], 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 [14].

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 [15]. 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 [16, 17]. 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 [18].

Notably, inclusion of patients with different mechanistic 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 [19, 20]. 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 [21]. 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 [22].

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 [23].

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 [24–26].

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 [27]. This analysis was run on just 725 patients and 1450 controls from UK Biobank, and contrasts with the 11 and 13 loci discovered using a GWAS approach respectively by 23andMe (16,500 patients/controls) [28] and COVID-19 HGI consortium (over 2,000,000 patient/controls) [29] in similar studies. Of the 68 genes that we reported, 48 were subsequently associated with the disease by other groups using methods including single-cell analysis and transcriptomic profiling (unpublished literature analysis—June 2022). This study went on to predict dutasteride as a drug repurposing candidate could be useful in significantly reducing symptom severity, the need for ICU and remission times in a subset of Covid-19 patients; findings which have been borne out by subsequent clinical trials [30].

Das *et al. Journal of Translational Medicine*   (2022) 20:598

Page 3 of 20

Materials and methods

We analyzed genotype data from 2382 patients reporting an ME/CFS diagnosis in the UK Biobank Pain Questionnaire [31] matched against 4764 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 (Additional file 1: Fig. S6). 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 Additional file 1: Table S5). To avoid potential confounding, controls meeting these criteria were matched by genetic ancestry and gender (Additional file 1: Fig. S7) against the cases in a 2:1 ratio (and this was repeated with a separate 4:1 ratio study).

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 Fig. 2 in Results section).
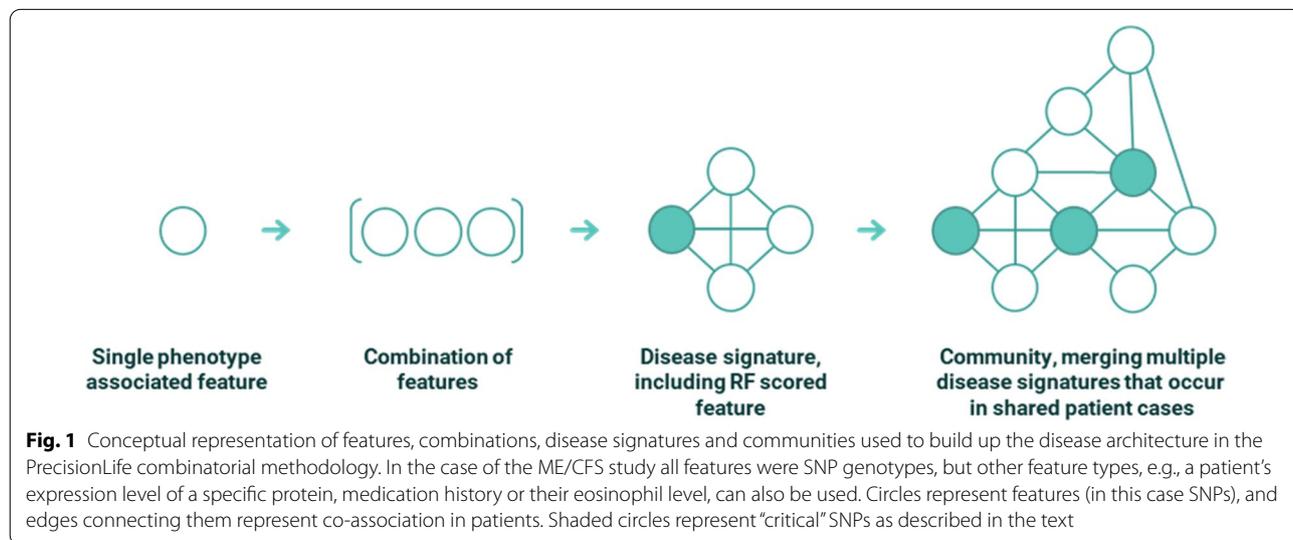
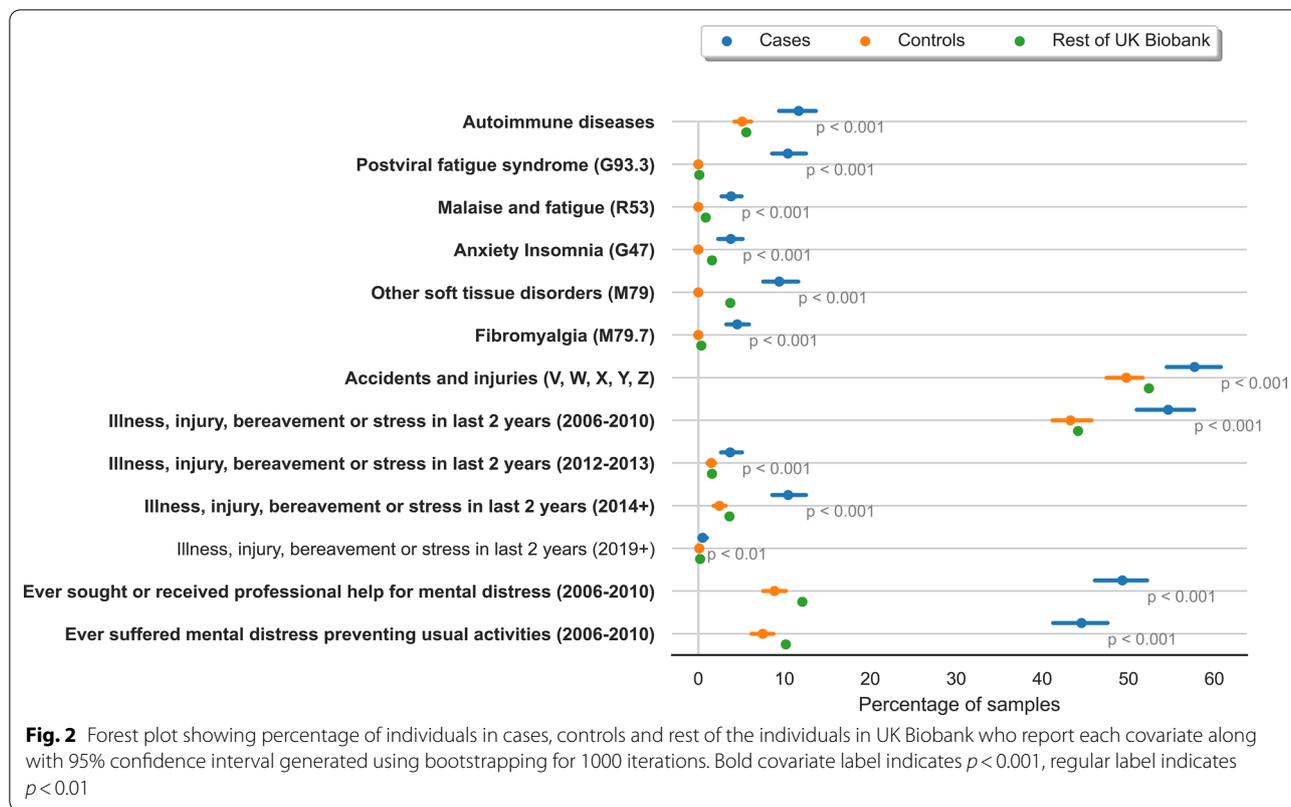After quality control (see "Genotype Quality Control" in Supplementary Data), the Pain Questionnaire dataset was comprised of 2382 ME/CFS cases, 4764 controls and 519,337 SNPs on autosomal chromosomes. Approximately 71% of cases (n = 1695) were women (Additional file 1: Fig. S7) versus the UK Biobank distribution of 54.4%. The age and body-mass index (BMI) distributions of cases and controls were similar (Additional file 1: Fig. S8).

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 (Fig. 1).

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.

Fig. 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. Circles represent features (in this case SNPs), and edges connecting them represent co-association in patients. Shaded circles represent "critical" SNPs as described in the text

Das *et al. Journal of Translational Medicine*   (2022) 20:598

Page 4 of 20

Fig. 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 1000 iterations. Bold covariate label indicates $p < 0.001$, regular label indicates $p < 0.01$

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 (Additional file 1: Tables S6 and S7). 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 1000 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 [32] are considered to be random and eliminated. Combinations passing these tests are reported as validated disease signatures.

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 fivefold 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

Das *et al. Journal of Translational Medicine*   (2022) 20:598

Page 5 of 20

druggability of targets is performed using a semantic knowledge graph (see "Functional Genomics Annotation" section).

For these studies, 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 $4 \times$ 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 1000 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 1000 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 considered significant.

In a second approach, we generated three new case populations from UK Biobank for fatigue-associated conditions and compared the results from these with the Pain Questionnaire cohort. These included a new CFS case population (Data-Field 20002, Coding 1482) reported during Verbal Interview, a post-viral syndrome (ICD10: G93.3) case population based on Hospital Episode Statistics, and a fibromyalgia (ICD10: M79.7) case population based on Hospital Episode Statistics.

The Verbal Interview case population comprised of 2270 individuals whose data had been analyzed in a recent GWAS study [33]. As the Verbal Interview cohort had 735 individuals in common with the Pain Questionnaire cohort (Additional file 1: Fig. S17), these overlapping cases were removed to create a disjoint Verbal Interview dataset (cases = 1273 and controls = 4137 after QC) of European ancestry with gender (and ancestry) matched controls. Similarly, disjoint post-viral syndrome dataset (cases = 510 and controls = 4763) and fibromyalgia dataset (cases = 1409 and controls = 4762 after QC) of European ancestry were generated.

The three disjoint datasets (Verbal Interview CFS, post-viral syndrome and fibromyalgia) were 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 these cohorts.

Two issues contribute to limit the degree of overlap that can be expected in a fully independent analysis of the disjoint cohorts. 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, within the two ME/CFS clinical diagnoses, the assignment by a GP of a specific diagnosis of "CFS" or "ME/CFS" or "post-viral syndrome" is highly variable and cannot be relied upon to distinguish these case populations in a clinically meaningful manner.

Due largely to the first sampling issue, it is likely that these three studies will differ significantly from the Pain Questionnaire study in the reported similarity of their genetic associations and clinical characteristics.

We therefore limited the search space for the analysis of the three disjoint cohorts 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) [34] 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 2 kb upstream or 0.5 kb 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 [35] and/or chromatin interaction data [36] (see Additional file 1: Table S12). 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.

Additionally, promoter capture Hi-C (pcHi-C) interactions that were significantly associated in brain tissues and blood cells were used to generate gene assignments. Due to the uncertainty about the relevant cells and

Das *et al. Journal of Translational Medicine*   (2022) 20:598

Page 6 of 20

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 disease studies) to avoid capturing any spurious associations from non-trait-related tissues. Genes that could be additionally mapped using only eQTL or HiC data from the 25 critical SNPs were observed and reported in Additional file 1: Table S12, 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 Additional file 1: Table S13). 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 [37] 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 (Fig. 2). The plot 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 1000 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.

Combinatorial analysis

Following quality control [38], the ME/CFS Pain Questionnaire cohort (2382 cases, 4764 controls) was used to perform a standard GWAS case–control association analysis using PLINK [39]. No SNPs were reported to be significant below a genome-wide significance threshold of $p < 5 \times 10^{-8}$ (Additional file 1: Fig. S9(a)).

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 1; Additional file 1: Table S7). None of the SNPs identified were observed to be in linkage disequilibrium (LD) with each other (Additional file 1: Fig. S10). 192 SNPs identified in the disease signatures were in non-coding regions of the genome and 9 SNPs (7 missense and 2 synonymous) were identified in the coding regions (Additional file 1: Fig. S11).

All SNPs were found in combinations with 3 or more SNPs, and so would not have been found using standard

Table 1 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%

Das *et al. Journal of Translational Medicine*   (2022) 20:598

Page 7 of 20

GWAS analysis methods (Additional file 1: Fig. S12). No single SNPs or SNP pairs were reported as significant by the method. The locations of the SNPs identified is shown in Additional file 1: Fig. S13.

As a negative validation check, runs using the same mining and validation parameters as used above were performed on 7146 random samples, comprising 2382 UK Biobank randomly sampled participants as 'cases' compared against 4764 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 1000 random permutation runs, and are therefore very unlikely to result from random chance. The odds ratios of the SNP combinations were found to be around 3.7 on average (Additional file 1: Fig. S12b). Additional file 1: Fig. S12c represents an example of a disease signature identified in this analysis containing five SNPs that were mapped to five genes.

Patient stratification

The disease architecture (Fig. 3), generated by clustering [40] 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.

There are 15 distinct communities of SNPs shown in the ME/CFS Pain Questionnaire disease architecture comprising between 142 to 744 of the 2382 cases (Additional file 1: Table S8). Odds ratios of the communities range from 2.16 to 4.47 and p-values from 10^{-10} to 10^{-72}. These share low (< 20%) patient overlap with each other (Additional file 1: Fig. S14), indicating they are distinct patient subgroups with different genetic drivers underlying their disease. Odds ratios of the communities range from 2.16 to 4.47 and p-values from 10^{-10} to 10^{-72}.

The analysis identified 25 critical disease associated SNPs (see the "Methods" and "Functional Genomics Annotation" sections) which are identified in multiple disease signatures. One of these SNPs—rs16947237, an intronic variant in *GPC5*—has previously been associated with self-reported chronic fatigue syndrome in a UK Biobank GWAS analysis [41].
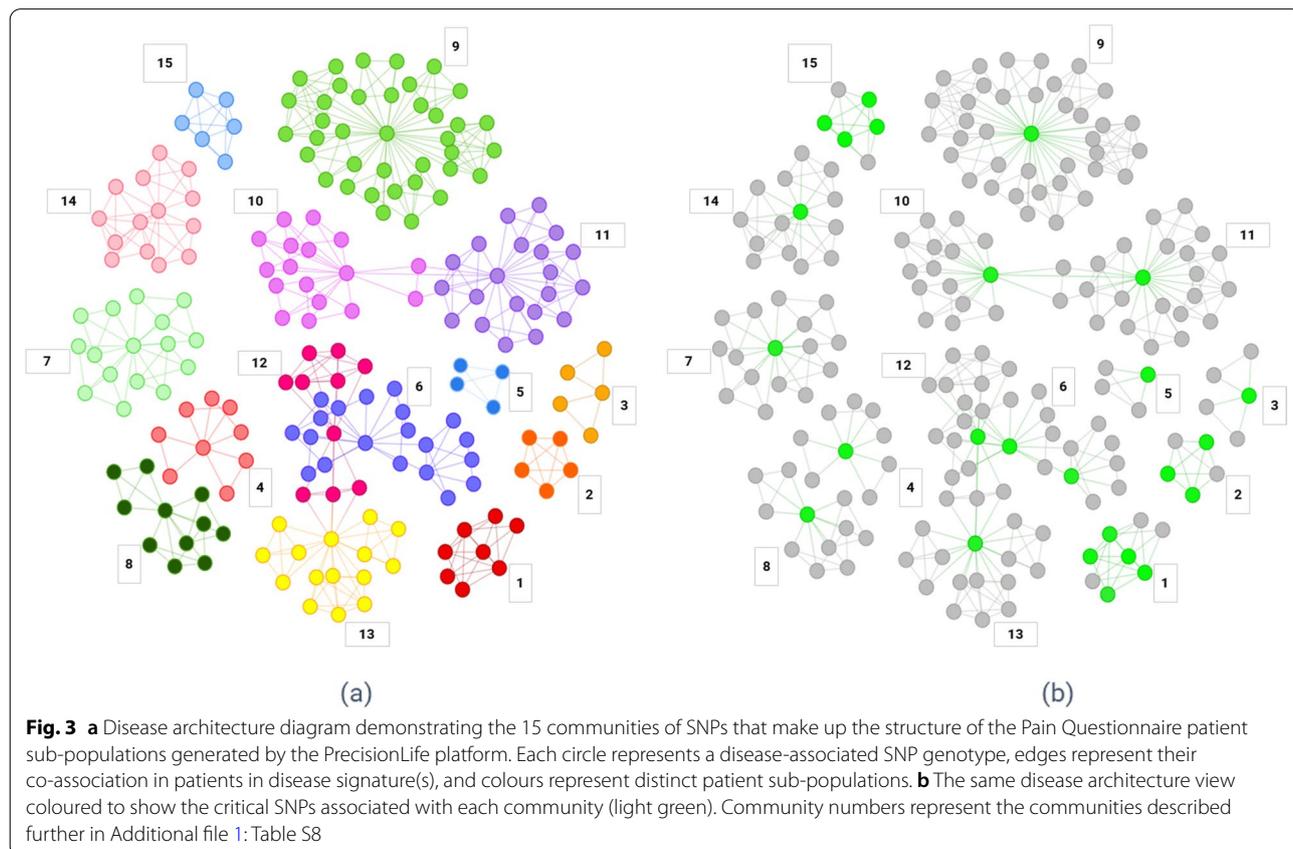

Fig. 3 **a** Disease architecture diagram demonstrating the 15 communities of SNPs that make up the structure of the Pain Questionnaire patient sub-populations generated by the PrecisionLife platform. Each circle represents a disease-associated SNP genotype, edges represent their co-association in patients 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). Community numbers represent the communities described further in Additional file 1: Table S8

Das *et al. Journal of Translational Medicine*   (2022) 20:598

Page 8 of 20

Table 2 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

Das *et al. Journal of Translational Medicine*   (2022) 20:598

Page 9 of 20

The critical SNPs (Additional file 1: Fig. S13) were mapped to 14 protein coding genes strongly associated with the ME/CFS Pain Questionnaire case population (Table 2). 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.

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 (Additional file 1: Table S12), when compared against the rest of the case population. These associations, however, did not reach statistical significance ($p < 0.05$) after multiple testing correction which may be due to the limited statistical power from the small dataset (Additional file 1: Tables S9 and S10).

Replication in three disjoint fatigue-associated cohorts

We tested for replication of the 25 critical SNPs associated with ME/CFS in the pain questionnaire in three additional disjoint cohorts of patients. Overall, 7 of the critical SNPs exhibited significant association with disease in at least one other cohort, with 2 replicated across multiple cohorts (Fig. 4).

Disjoint CFS (Verbal Interview) cohort

We analyzed the disjoint UK Biobank CFS Verbal Interview cohort (1273 cases, excluding individuals that were common to the Pain Questionnaire cohort, and 4137 controls) using GWAS and combinatorial analysis.

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 (Additional file 1: Fig. S9b), 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 (Additional file 1: Fig. S9a). 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.

Comparison of the results validating the 199 SNPs from the Pain Questionnaire cohort in the Verbal Interview cohort showed that five (rs2304725, rs2904106, rs9444564, rs10420798 and rs11695478) of the 25 critical SNPs 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.

Disjoint post-viral syndrome cohort

Three critical SNPs (rs56218501/Affx-16805420, rs12530627 and rs2499908) identified in the Pain Questionnaire cohort were replicated in the fibromyalgia cohort analysis. Two of these critical SNPs mapped to the genes SULF2 (rs56218501/Affx-16805420) and USP6NL (rs2499908).

Disjoint fibromyalgia cohort

Two critical SNPs (rs10420798 and rs2499908) and the gene USP6NL (mapped to rs2499908) identified in the Pain Questionnaire cohort were replicated in the fibromyalgia cohort analysis. This suggests that these two critical SNPs and the gene USP6NL are particularly strongly associated with fibromyalgia cases.

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 [109] 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
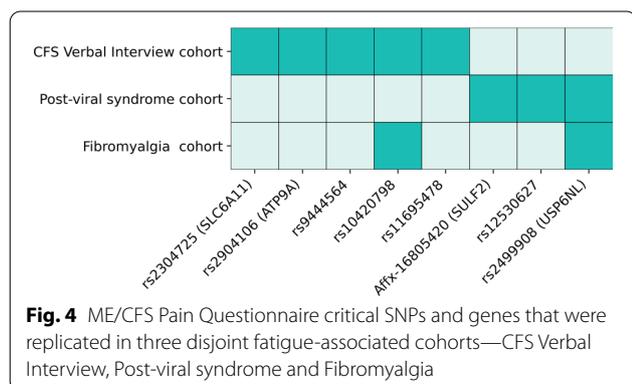

Fig. 4 ME/CFS Pain Questionnaire critical SNPs and genes that were replicated in three disjoint fatigue-associated cohorts—CFS Verbal Interview, Post-viral syndrome and Fibromyalgia

Das *et al. Journal of Translational Medicine*   (2022) 20:598

Page 10 of 20

Table 3 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 [42]
ATP9A	1	Metabolic	Exhibits P4-ATPase activity in beta cells and *ATP9A* knockdown results in reduction in glucose-stimulated insulin release [43]
		Autoimmune	Splicing region in *ATP9A* associated with multiple sclerosis risk in whole-exome sequencing analysis [44]
KCNB1	1	Metabolic	Regulates insulin exocytosis from pancreatic beta cells [45]
CLOCK	2	Sleep	Key regulator of circadian rhythm [46]
		Metabolic	Disruption to circadian clock results in mitochondrial dysfunction and insulin resistance [47, 48]
SLC15A4	3	Autoimmune	SLC15A4 mediates type 1 interferon production and generation of autoantibodies in systemic lupus erythematosus (SLE) [49]
		Metabolic	Knock down of *SLC15A4* decreases mitochondrial membrane potential and impairs AMPK activation [50, 51]
TMEM232	6	Autoimmune	The promoter region of *TMEM232* is differentially methylated in a study of multiple sclerosis patients [52]
GPC5	7	Autoimmune	Polymorphisms in GPC5 have been associated with multiple sclerosis risk in GWAS [53, 54]
PHACTR2	8	Autoimmune	Identified as a multiple sclerosis risk gene in a GWA analysis [55]
AKAP1	9	Metabolic	AKAP1 is an AMPK substrate that regulates mitochondrial respiration [56]
		Viral/bacterial susceptibility	*AKAP1-/-* mice exhibit increased inflammation and immune cell infiltration in lungs [57]
USP6NL	12	Viral/bacterial susceptibility	Plays a role in HSV-1 virion assembly [58]
CDON	13	Metabolic	Cdon depletion results in impaired muscle regeneration and increased cell stress [59]
		Autoimmune	Cdon is involved in differentiation and myelination of oligodendrocytes [60]
		Viral/bacterial susceptibility	GWAS indicates *CDON* plays a role in complicated *Staphylococcus aureus* bacteremia [61]
INSR	14	Metabolic	Receptor tyrosine kinase which mediates insulin signalling [62]
SULF2	15	Viral/bacterial susceptibility	Genome-wide environmental interaction study indicates an association between a locus in *SULF2* and HSV-1 seropositivity and depression [63]
		Metabolic	Overexpression of *SULF2* observed in subjects with type 2 diabetes and obesity [64]
SLC6A11	15	Sleep	Associated with depression, sleep disturbance and neuropathic pain in mouse models [65, 66]
		Autoimmune	Expression decreased in the hippocampus of mouse model of multiple sclerosis (EAE) [67]

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 3).

Furthermore, critical SNPs found in the same disease signature and/or patient community may be mapped to genes with shared biological functions or pathways. Pathway enrichment analysis [68] was performed on each of the 199 disease signatures using the genes that were associated to any constituent SNP (see section "Pathway Enrichment Analysis" in Supplementary Data). Analysis of the enrichment results indicated that two large communities containing multiple disease signatures and critical SNPs—Community 1 and Community 15 (Table 4)—may be implicated in common biological processes (Additional file 1: Table S14).

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 co-occurring in six common disease signatures (Additional file 1: Table S9), which shows significant enrichment linked to regulation of exocytosis and negative regulation of secretion by cell (GO annotations, Additional file 1: Table S14). 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 4) [36, 38]. This could suggest a 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 co-occurring in two disease signatures; rs2304725, a synonymous variant in *SLC6A11*, and rs56218501/ Affx-16805420, a missense variant in *SULF2* (Additional file 1: Table S9). This disease signature shows enrichment for GABA synthesis and release and synaptic vesicle cycle (Reactome and KEGG annotations, Additional file 1: Table S14). This

Table 4 Genes identified in the Pain Questionnaire cohort with their tractability as drug targets using annotations from OpenTargets [109]

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

enrichment can be supported by further literature evidence that indicates the association of these genes with CNS-related pathways (Table 4) [56, 58, 59].

Many of the identified patient communities 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 (Fig. 5). 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.

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 [69]. There have been reports of shared pathophysiological, clinical, and transcriptomic features between viral and/or bacterial diseases and ME/CFS [70, 71].

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 4).

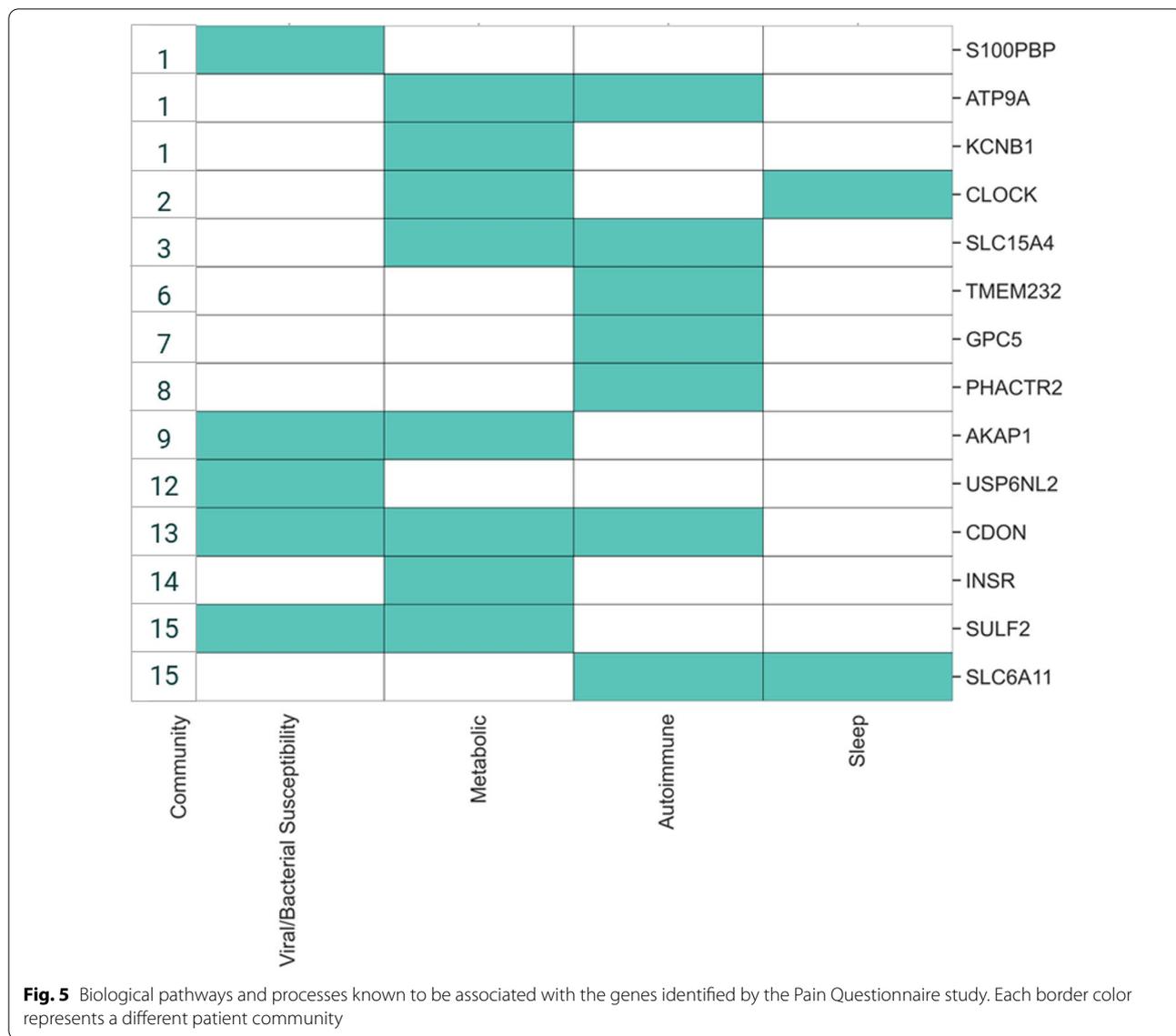These may represent a subset of ME/CFS patients with increased susceptibility to infection, or differential response to infection that leads to ineffectual viral clearance. We therefore evaluated the clinical records of the ME/CFS case population included in this study to identify any evidence of prior 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) and incomplete 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 proinflammatory cytokines and higher prevalence in females, with as many as 60% of ME/CFS patients also reported to be diagnosed with an autoimmune disease [72, 73]. 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 (Fig. 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

Fig. 5 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

and autoimmune development [74]. 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 [75–77].

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 [78]. 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 [51] and *SLC15A4* knock down results in decreased mitochondrial function under cell stress [50]. 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 [53, 79, 80]. 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 4). 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 [81].

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 [82].

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 [83]. 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 [56]. 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 [84].

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 [85], which have been shown to be altered in patients with chronic fatigue and in the skeletal muscle of fatigue-associated conditions [86–88]. ATP9A is also expressed in pancreatic beta cells and has a role in driving glucose-stimulated insulin release [43]. Moreover, a variant in *ATP9A* has been associated with multiple sclerosis in a homozygosity haplotype analysis [44].

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 [89], which is hypothesized to be as a results of insulin resistance and ischemia–reperfusion damage in skeletal muscles of patients with ME/CFS [90].

These 348 ME/CFS patients also presented with relatively higher blood levels of lactate from the UK Biobank NMR metabolomics data ($p < 0.017$, Additional file 1: Table S11) 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 [91, 92]. Dysregulation of insulin and lactate in ME/CFS patients may also have an impact on mitochondrial function, decreasing mitochondrial size and respiratory function [93].

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 ($p = 0.003$, Additional file 1: Table S9) 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 [94], in addition to sleep disturbance, juvenile stress and depression in animal models [66, 95–97]. Furthermore, patients with SNP combinations including those in *SLC6A11* show increased levels of phenylalanine ($p = 0.022$) in the metabolomics data compared to other ME/CFS subgroups identified in our analysis (although this is not significant after multiple testing correction). 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 (Additional file 1: Fig. S16). Sulfatase 2 is crucial for brain development, contributing to processes such as neurite outgrowth and responsiveness to growth factors [98–100], and there is an association between *SULF2* variants and HSV-1 and depression risk, and also with malaise and fatigue in UK Biobank studies [63, 101]. The directionality of the specific observed SULF2 association, i.e. whether this is a 'risk' allele or a 'protective' allele, is not however clear from this study or the literature. Although there is a known association with sex hormone globulin levels, we did not find any difference in male:female distribution for this community.

Das *et al. Journal of Translational Medicine*   (2022) 20:598

Page 14 of 20

CDON (cell adhesion associated, oncogene regulated) encodes a cell surface receptor that is highly involved in muscle regeneration [63]. In muscle cells, depletion of CDON results in impaired muscle regeneration and senescence, as well as increased cell stress [102]. However, CDON has also been associated with complicated bacteremia [61] and development of midbrain dopamine pathways [103]. 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 [82]. 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 [47, 104, 105]. Furthermore, transcriptomic analysis of peripheral blood mononuclear cells indicated that several genes involved in circadian rhythm were elevated in ME/CFS patients [106].

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 [107]. 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 this combinatorial analysis of the UK Biobank population may 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 [108] 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. Before we can use such in vitro/in vivo tools to evaluate the effects of modulating novel targets either pharmacologically or via direct genetic manipulation, further work on identifying cell types, phenotypic readouts and animal models will have to be undertaken.

Each gene identified in this study was nonetheless evaluated for drug tractability (Table 4), 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, especially in the absence of in vitro/in vivo tools, if their safety and efficacy can be demonstrated.

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 [51, 63, 64, 68] and show a level of overlap with diseases sharing similar symptoms, such as MS [110] and long Covid [111, 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 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, recovery or relapse.

While we have tried to replicate the analysis and results between two different ME/CFS 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 in these small datasets. 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 [9], perhaps with an earlier secondary peak [113], whereas the average age of the UK Biobank population is 56 years [114] and the population has a selective participation bias to 'healthy volunteers' [115]. 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 [31]. 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 [81], and both can have a viral trigger such as Epstein-Barr virus (EBV) [4, 116]. 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 [11, 117, 118].

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 [119]. 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) 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

Das *et al. Journal of Translational Medicine*   (2022) 20:598

Page 16 of 20

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 future publication.

Conclusion/future perspectives

The hypothesis-free combinatorial analytics approach implemented in the PrecisionLife platform identified 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 [120, 121], 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 towards 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 [122], 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.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-022-03815-8.

> **Additional file 1**: Supplemental data from the analyses described in the paper.

Author 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.

Availability of data and materials

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.

Das *et al. Journal of Translational Medicine*   (2022) 20:598

Page 17 of 20

Declarations

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;12(1):2157. https://doi.org/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;9:91. https://doi.org/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;9(3):91. https://doi.org/10.3390/diagnostics9030091.
4. Lorusso L, Mikhaylova SV, Capelli E, Ferrari D, Ngonga GK, Ricevuti G. Immunological aspects of chronic fatigue syndrome. Autoimmun Rev. 2009;8(4):287–91. https://doi.org/10.1016/j.autrev.2008.08.003.
5. 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;12: 656797. https://doi.org/10.3389/fimmu.2021.656797.
6. Rasa S, Nora-Krukle Z, Henning N, Eliassen E, Shikova E, Harrer T, Scheibenbogen C, Murovska M, Prusty BK. 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. https://doi.org/10.1186/s12967-018-1644-y.
7. Hickie I, Davenport T, Wakefield D, Vollmer-Conna U, Cameron B, Vernon SD, Reeves WC, Lloyd A. Post-infective and chronic fatigue syndromes precipitated by viral and non-viral pathogens: prospective cohort study. BMJ. 2006. https://doi.org/10.1136/bmj.38933.585764.AE.
8. Katz BZ, Shiraishi Y, Mears CJ, Binns HS, Taylor R. Chronic fatigue syndrome after infectious mononucleosis in adolescents. Pediatrics. 2009;124:189–93. https://doi.org/10.1542/peds.2008-1879].
9. Chu L, Valencia IJ, Garvet DW, Montoya JG. Onset patterns and course of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. Front Pediatr. 2019;7:12. https://doi.org/10.3389/fped.2019.00012.
10. 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;18(20):10614. https://doi.org/10.3390/ijerph182010614.
11. 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;8:20499361211009384. https://doi.org/10.1177/20499361211009385.
12. Ueland M, Hajdarevic R, Mella O, Strand EB, Sosa DD, Saugstad OD, Fluge Ø, Lie BA, Viken MK. No replication of previously reported association with genetic variants in the T cell receptor alpha (TRA) locus for

13. myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). Transl Psychiatry. 2022;12(1):277. https://doi.org/10.1038/s41398-022-02046-1.
13. Albright F, Light K, Light A, Bateman L, Cannon-Albright LA. Evidence for a heritable predisposition to Chronic Fatigue Syndrome. BMC Neurol. 2011;11:62. https://doi.org/10.1186/1471-2377-11-62.
14. Dibble JJ, McGrath SJ, Ponting CP. Genetic risk factors of ME/CFS: a critical review. Hum Mol Genet. 2020;29(R1):R117–24. https://doi.org/10.1093/hmg/ddaa169.
15. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019;20(8):467–84. https://doi.org/10.1038/s41576-019-0127-1.
16. Horesh Bergquist S, Lobelo F. The limits and potential future applications of personalized medicine to prevent complex chronic disease. Public Health Rep. 2018;133(5):519–22. https://doi.org/10.1177/0033354918781568.
17. Strawbridge RJ, van Zuydam NR. Shared genetic contribution of type 2 diabetes and cardiovascular disease: implications for prognosis and treatment. Curr Diab Rep. 2018;18(8):59. https://doi.org/10.1007/s11892-018-1021-5.
18. Abell NS, DeGorter MK, Gloudemans MJ, Greenwald E, Smith KS, He Z, Montgomery SB. Multiple causal variants underlie genetic associations in humans. Science. 2022;375(6586):1247–54. https://doi.org/10.1126/science.abj5117.
19. Chang YC, Wu JT, Hong MY, Tung YA, Hsieh PH, Yee SW, Giacomini KM, Oyang YJ, Chen CY; Alzheimer's Disease Neuroimaging Initiative. GenEpi: gene-based epistasis discovery using machine learning. BMC Bioinform. 2020;21(1):68. https://doi.org/10.1186/s12859-020-3368-2.
20. Moore JH, Williams SM. Epistasis and its implications for personal genetics. Am J Hum Genet. 2009;85(3):309–20. https://doi.org/10.1016/j.ajhg.2009.08.006.
21. Wray NR, Wijmenga C, Sullivan PF, Yang J, Visscher PM. Common disease is more complex than implied by the core gene omnigenic model. Cell. 2018;173(7):1573–80. https://doi.org/10.1016/j.cell.2018.05.051.
22. Visscher PM, Yengo L, Cox NJ, Wray NR. Discovery and implications of polygenicity of common diseases. Science. 2021;373(6562):1468–73. https://doi.org/10.1126/science.abi8206.
23. Gardner, S. Combinatorial analytics: an essential tool for the delivery of precision medicine and precision agriculture. Artif Intell Life Sci. 2021;1:100003. https://doi.org/10.1016/j.ailsci.2021.100003
24. Koefoed P, Andreassen OA, Bennike B, Dam H, Djurovic S, Hansen T, Jorgensen MB, Kessing LV, Melle I, Møller GL, et al. Combinations of SNPs related to signal transduction in bipolar disorder. PLoS ONE. 2011;6: e23812. https://doi.org/10.1371/journal.pone.0023812.
25. Das S, Pearson M, Taylor K, Bouchet V, Møller GL, Hall TO, Strivens M, Tzeng KT, Gardner S. Combinatorial analysis of phenotypic and clinical risk factors associated with hospitalized COVID-19 patients. Front Digit Health. 2021;3: 660809. https://doi.org/10.3389/fdgth.2021.660809.
26. Taylor K, Das S, Pearson M, Kozubek J, Strivens M, Gardner S. Systematic drug repurposing to enable precision medicine: a case study in breast cancer. Digit Med. 2019;5:180. https://doi.org/10.4103/digm.digm_28_19.
27. Taylor K, Das S, Pearson M, Kozubek J, Pawlowski M, Jensen CE, Skowron Z, Møller GL, Strivens M, Gardner S. Analysis of genetic host response risk factors in severe COVID-19 patients. Preprint at medRxiv. 2020. https://doi.org/10.1101/2020.06.17.20134015.
28. 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;53(6):801–8. https://doi.org/10.1038/s41588-021-00854-7.
29. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. Nature. 2021;600(7889):472–7. https://doi.org/10.1038/s41586-021-03767-x.
30. Cadegiani FA, McCoy J, Gustavo Wambier C, Goren A. Early antiandrogen therapy with dutasteride reduces viral shedding, inflammatory responses, and time-to-remission in males with COVID-19: a randomized, double-blind, placebo-controlled interventional trial (EAT-DUTA AndroCoV Trial - Biochemical). Cureus. 2021;13(2): e13047. https://doi.org/10.7759/cureus.13047.
31. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, et al. The UK Biobank resource

Das *et al. Journal of Translational Medicine*   (2022) 20:598

Page 18 of 20

with deep phenotyping and genomic data. Nature. 2018;562:203–9. https://doi.org/10.1038/s41586-018-0579-z.

32. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple hypothesis testing. J R Stat Soc B. 1995;57:289–300.

33. 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;102:362–9. https://doi.org/10.1016/j.bbi.2022.03.010.

34. 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;49(D1):D884–91. https://doi.org/10.1093/nar/gkaa942.

35. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369(6509):1318–30.

36. 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. Nat Genet. 2019;51(10):1442–9.

37. 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. Nat Rev Drug Discov. 2018;17(3):167–81.

38. Jin Y, Schäffer AA, Sherry ST, Feolo M. Quickly identifying identical and closely related subjects in large databases using genotype data. PLoS ONE. 2017;12(6): e0179106. https://doi.org/10.1371/journal.pone.0179106.

39. 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. Am J Hum Genet. 2007;81(3):559–75.

40. 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;12(1):8248. https://doi.org/10.1038/s41598-022-11987-y.

41. Neale lab, UK Biobank v2 (2018). https://www.nealelab.is/uk-biobank

42. Gardinassi LG. A cross-study biomarker signature of human bronchial epithelial cells infected with respiratory syncytial virus. Adv Virol. 2016;2016:3605302. https://doi.org/10.1155/2016/3605302.

43. Ansari IU, Longacre MJ, Paulusma CC, Stoker SW, Kendrick MA, Mac-Donald 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;290(38):23110–23. https://doi.org/10.1074/jbc.M115.655027.

44. 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;43(3):1778–93. https://doi.org/10.3390/cimb43030125.

45. 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;344(2):407–16. https://doi.org/10.1124/jpet.112.199083

46. Huang W, Ramsey KM, Marcheva B, Bass J. Circadian rhythms, sleep, and metabolism. J Clin Invest. 2011;121(6):2133–41. https://doi.org/10.1172/JCI46043.

47. de Goede P, Wefers J, Brombacher EC, Schrauwen P, Kalsbeek A. Circadian rhythms in mitochondrial respiration. J Mol Endocrinol. 2018;60(3):R115–30. https://doi.org/10.1530/JME-17-0196.

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

49. 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;41(3):375–88. https://doi.org/10.1016/j.immuni.2014.08.011.

50. 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;33(7):399–406. https://doi.org/10.1093/intimm/dxab006.

51. 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;118(33): e2100295118. https://doi.org/10.1073/pnas.2100295118.

52. 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;10(1):2094. https://doi.org/10.1038/s41467-019-09984-3.

53. 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;64(2):230–4. https://doi.org/10.1016/j.advms.2018.12.004.

54. 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;8(10): e75416. https://doi.org/10.1371/journal.pone.0075416.PMID:24130709.

55. 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;11:49. https://doi.org/10.1186/1471-2156-11-49.

56. 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;71(3):507–17. https://doi.org/10.1161/HYPERTENSIONAHA.117.10185.

57. 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;314(5):L860–70. https://doi.org/10.1152/ajplung.00365.2017.

58. 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;85(16):8012–21. https://doi.org/10.1128/JVI.00500-11.

59. 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;11(4):1089-1103. doi: 10.1002/jcsm.12563. J Cachexia Sarcopenia Muscle. 2020;11(5):1381.

60. Wang LC, Almazan G. Cdon, a cell surface protein, mediates oligodendrocyte differentiation and myelination. Glia. 2016;64(6):1021–33. https://doi.org/10.1002/glia.22980.

61. Shukla SK, Rose W, Schrodi SJ. Complex host genetic susceptibility to Staphylococcus aureus infections. Trends Microbiol. 2015;23(9):529–36. https://doi.org/10.1016/j.tim.2015.05.008.

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

63. 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;10(2): e108. https://doi.org/10.1002/ctm2.108.

64. 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 predisposes to fasting and postprandial triglycerides in patients with obesity and type 2 diabetes mellitus. Obesity (Silver Spring). 2014;22(5):1309–16. https://doi.org/10.1002/oby.20682.

Das *et al. Journal of Translational Medicine*   (2022) 20:598

Page 19 of 20

65. 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;152(6):1358–72. https://doi.org/10.1016/j.pain.2011.02.016.

66. 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;68(6):235–47. https://doi.org/10.1002/syn.21733.

67. 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;395:89–100. https://doi.org/10.1016/j.neuroscience.2018.11.003.

68. 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 Res. 2019;47(W1):W191–8.

69. 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;333(7568):575. https://doi.org/10.1136/bmj.38933.585764.AE

70. 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;18(1):448. https://doi.org/10.1186/s12967-020-02585-5.

71. 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;57(5):418. https://doi.org/10.3390/medicina57050418.

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

73. 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;10(20):4786. https://doi.org/10.3390/jcm10204786.

74. 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;16(1):268. https://doi.org/10.1186/s12967-018-1644-y

75. 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;29(1):19–36. https://doi.org/10.1007/s11011-013-9435-x.

76. 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;20(8): 102867. https://doi.org/10.1016/j.autrev.2021.102867.

77. Clayton SA, MacDonald L, Kurowska-Stolarska M, Clark AR. Mitochondria as key players in the pathogenesis and treatment of rheumatoid arthritis. Front Immunol. 2021;12: 673916. https://doi.org/10.3389/fimmu.2021.673916.

78. 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;21(9):994–9. https://doi.org/10.1038/ejhg.2012.277.

79. 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;335(1–2):82–8. https://doi.org/10.1016/j.jns.2013.08.031.

80. 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;11(4):343–50. https://doi.org/10.1038/gene.2009.81.

81. 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;17(1):117. https://doi.org/10.1186/s12883-017-0896-0.

82. 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;12(10): e0186802. https://doi.org/10.1371/journal.pone.0186802.

83. 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;9(2):298. https://doi.org/10.3390/cells9020298.

84. Yoshinaka T, Kosako H, Yoshizumi T, Furukawa R, Hirano Y, Kuge O, Tamada T, Koshiba T. Structural basis of mitochondrial scaffolds by prohibition complexes: insight into a role of the coiled-coil region. iScience. 2019;19:1065–1078. https://doi.org/10.1016/j.isci.2019.08.056.

85. 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;2020:8865499. https://doi.org/10.1155/2020/8865499.

86. Nikolova-Karakashian MN, Reid MB. Sphingolipid metabolism, oxidant signaling, and contractile function of skeletal muscle. Antioxid Redox Signal. 2011;15(9):2501–17. https://doi.org/10.1089/ars.2011.3940.

87. 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]. 2022:2021.06.14.21258895. doi: https://doi.org/10.1101/2021.06.14.21258895

88. 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;8(1):10056. https://doi.org/10.1038/s41598-018-28477-9.

89. 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;41(5):567–73. https://doi.org/10.1016/s0006-3223(96)00074-1.

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

91. 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;283(2):E233–40. https://doi.org/10.1152/ajpendo.00557.2001.

92. 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;21(13):4715. https://doi.org/10.3390/ijms21134715.

93. 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;6:22788. https://doi.org/10.1038/srep22788.

94. 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;13(1):83. https://doi.org/10.1186/s12974-016-0549-z.

95. 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;152(6):1358–72. https://doi.org/10.1016/j.pain.2011.02.016.

Das *et al. Journal of Translational Medicine*   (2022) 20:598

Page 20 of 20

96. 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;34(9):1584–93. https://doi.org/10.1007/s11064-009-9947-2.

97. 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;64(6):911–22. https://doi.org/10.1002/glia.22970.

98. 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;10(10): e0139853. https://doi.org/10.1371/journal.pone.0139853.

99. 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;13(11–12):4505–21. https://doi.org/10.1111/j.1582-4934.2008.00558.x.

100. 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;263:150–60. https://doi.org/10.1016/j.expneurol.2014.10.011.

101. Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet. 2018;50(9):1335–41. https://doi.org/10.1038/s41588-018-0184-y.

102. 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;11(4):1089–103. https://doi.org/10.1002/jcsm.12563.

103. 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;3(3). https://doi.org/10.1523/ENEURO.0009-16.2016

104. 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;27(3):657-666.e5. https://doi.org/10.1016/j.cmet.2018.01.011.

105. Oosterman JE, Wopereis S, Kalsbeek A. The circadian clock, shift work, and tissue-specific insulin resistance. Endocrinology. 2020;161(12):bqaa180. https://doi.org/10.1210/endocr/bqaa180.

106. 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. https://doi.org/10.1177/2058738418820402.

107. Natelson BH. Myalgic encephalomyelitis/chronic fatigue syndrome and fibromyalgia: definitions, similarities, and differences. Clin Ther. 2019;41(4):612–8. https://doi.org/10.1016/j.clinthera.2018.12.016.

108. Gatto N, Dos Santos SC, 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;20(1): e13281. https://doi.org/10.1111/acel.13281.

109. 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;49(D1):D1302–10. https://doi.org/10.1093/nar/gkaa1027.

110. 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 Biomed. 2017;2(2).

111. 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. J Adv Res. 2021.

112. Marshall-Gradisnik S, Eaton-Fitch N. Understanding myalgic encephalomyelitis. Science. 2022;377(6611):1150–1.

113. 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;12:167. https://doi.org/10.1186/s12916-014-0167-5.

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

115. 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;47(4):1207–16. https://doi.org/10.1093/ije/dyy060.

116. 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;375(6578):296–301. https://doi.org/10.1126/science.abj8222.

117. Komaroff AL, Bateman L. Will COVID-19 lead to myalgic encephalomyelitis/chronic fatigue syndrome? Front Med (Lausanne). 2021;7: 606824. https://doi.org/10.3389/fmed.2020.606824.

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

119. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. Nature. 2021;600:472–7.

120. https://www.decodeme.org.uk/

121. Devereux-Cooke A, Leary S, McGrath SJ, Northwood E, Redshaw A, Shepherd C, Stacey P, Tripp C, Wilson J, Mar M, Boobyer D, Bromiley S, Chowdhury S, Dransfield C, Almas M, Almelid Ø, Buchanan D, Garcia D, Ireland J, Kerr SM, Lewis I, McDowall E, Migdal M, Murray P, Perry D, Ponting CP, Vitart V, Wolfe JC. DecodeME: community recruitment for a large genetics study of myalgic encephalomyelitis/chronic fatigue syndrome. BMC Neurol. 2022;22(1):269. https://doi.org/10.1186/s12883-022-02763-6.

122. 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;3(6): 100496. https://doi.org/10.1016/j.patter.2022.100496.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.