

Article

Not peer-reviewed version

---

# Genetic Insights into Circulating Complement Proteins in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: A Potential Inflammatory Subgroup

---

[Jessica Maya](#)\*, [Elizabeth R Unger](#), [Jin-Mann S. Lin](#), [Mangalathu S Rajeevan](#)\*

Posted Date: 19 December 2025

doi: 10.20944/preprints202512.1773.v1

Keywords: ME/CFS; complement system; pQTLs; myalgic encephalomyelitis/chronic fatigue syndrome; heterogeneity



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Genetic Insights into Circulating Complement Proteins in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: A Potential Inflammatory Subgroup

Jessica Maya \*, Elizabeth R. Unger, Jin-Mann S. Lin and Mangalathu S. Rajeevan \*

Division of High-Consequence Pathogens & Pathology, Centers for Disease Control & Prevention, Atlanta, GA 30329, USA

\* Correspondence: xpa0@cdc.gov (J.M.); msrajeevan53@gmail.com (M.S.R.)

## Abstract

Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) is a debilitating multi-system illness with heterogeneity that complicates identifying the pathophysiology, biomarkers, and therapeutic targets. Evidence indicates the importance of immune dysregulation, including the complement system, in ME/CFS. This study investigates the contribution of genetic drivers to potential dysregulation of the complement pathway in ME/CFS. We used protein quantitative trait loci (pQTL) analyses, adjusted for covariates using linear and logistic regression, to identify genetic variants significantly associated with plasma complement protein levels in a study sample identified from the general population (50 ME/CFS and 121 non-fatigued). ME/CFS patients carrying certain pQTLs exhibited dysregulation of the alternative complement pathway, which defined an inflammatory subgroup with a high C3/low Bb profile and established a genetic link to dysregulation of the alternative complement pathway. Six of the significant pQTLs were also found associated with fatigue-related phenotypes in the UK biobank, 4 of which were complement-associated, providing some validation in an independent population. Our findings highlight a mechanism by which risk alleles contribute to ME/CFS heterogeneity, providing evidence of a genetic basis for complement dysregulation in a subgroup. This approach could identify pathway-focused subgroups in ME/CFS and similar illnesses to inform personalized approaches to diagnosis and treatments.

**Keywords:** ME/CFS; complement system; pQTLs; myalgic encephalomyelitis/chronic fatigue syndrome; heterogeneity

---

## 1. Introduction

Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) is a debilitating multi-system illness characterized by substantial functional impairment accompanied by profound fatigue, post-exertional malaise, unrefreshing sleep, cognitive dysfunction, orthostatic intolerance, and a wide litany of additional symptoms that significantly reduce quality of life [1]. Its etiology remains elusive, with substantial heterogeneity complicating efforts to identify reliable biomarkers and therapeutic targets [2]; however, understanding the molecular basis for genotype/phenotype associations can greatly enhance our ability to tailor diagnostic and treatment strategies. While this approach is considered useful, it has yet to be investigated for many complex and heterogeneous diseases, including ME/CFS. Recently, we and others have tested whole genome or pathway focused genetic markers for their association with ME/CFS [3–14]. Many results from these earlier genetic association studies have not been replicated, possibly due to small sample size, lack of consistency in case

ascertainment, heterogeneity at the level of genetics, disease time-course, co-morbid illnesses, and their combinatorial interaction with environment, diet, medications, and microbial environments.

There has been growing interest in exploring the relationship between disease-associated genetic variants and the levels of RNA and proteins, known as expression/protein quantitative trait loci (eQTLs/pQTLs) respectively. These intermediate phenotypes bridge the gap between genotype and clinical presentation, offering a powerful framework for understanding complex diseases and informing biomarker and therapeutic target discovery. While pQTL studies have been applied to cardiovascular disease, a variety of autoimmune disorders, and proteins in the human liver [15–17], their application to ME/CFS remains limited.

Our group previously identified several genetic variants within the complement pathway that associated with ME/CFS using a focused analysis of immune and inflammation-related genes [3]. The complement system is a critical component of innate immunity, consisting of plasma proteins that defend against infection and mediate inflammation through three activation pathways – classical, alternative, and lectin – all converging at C3 cleavage and tightly regulated to prevent chronic inflammation and immune-mediated damage [18,19]. Notably, our prior study identified two non-synonymous single nucleotide polymorphisms (SNPs) in complement-related genes associated with ME/CFS: rs4151667 in *Complement Factor B (CFB)* and rs1061170 in *Complement Factor H (CFH)*, both of which regulate the alternative pathway of complement activation. The SNP rs4151667 in *CFB* results in a leucine-to-histidine substitution at position 9 (L9H) and is in strong linkage disequilibrium (LD) with rs9332739 in *C2*. Similarly, rs1061170 in *CFH* causes a tyrosine-to-histidine substitution at position 404 (Y402H) and is in high LD with three other SNPs in *CFH* (rs1061147, rs7529589, and rs1080155). Both rs9332739 and rs4151667, located in the paralogous *C2* and *CFB* genes, were significantly associated with ME/CFS in allele and haplotype analyses. Likewise, the major alleles of all four SNPs in *CFH* were associated with ME/CFS, further implicating the complement pathway in disease susceptibility. Interestingly, previous studies have reported complement dysregulation in ME/CFS – including elevated C4a responses, unique correlations between C3 levels and other circulating inflammatory proteins, and altered lectin and classical pathway components in plasma, cerebrospinal fluid, and extracellular vesicles [20–26]. Integrating genetic and proteomic data has the potential to better understand disease susceptibility, heterogeneity, and progression.

It should be noted that *CFH* and *CFB* have opposing roles in the alternative complement pathway, with *CFB* promoting and *CFH* inhibiting the decay of C3 convertase [27]. Based on this observation, we hypothesize that alternative pathway components, consisting of precursor proteins (C3 and Factor B), their breakdown products (C3a, C5a, Bb), negative regulator (Factor H), and marker of terminal pathway of complement activation (SC5b-9), would be associated with ME/CFS, and that this association may be driven by the observed genetic variation in the complement proteins. Interestingly, ME/CFS-associated missense variants in these genes are also linked to age-related macular degeneration (AMD) but with inverted risk alleles [28,29]. As these alleles associate with reduced complement activation and protection in AMD [30–32], we hypothesize that reduced complement activity contributes to disease risk in ME/CFS. These risk alleles could enable identification of a subgroup of ME/CFS with complement dysregulation.

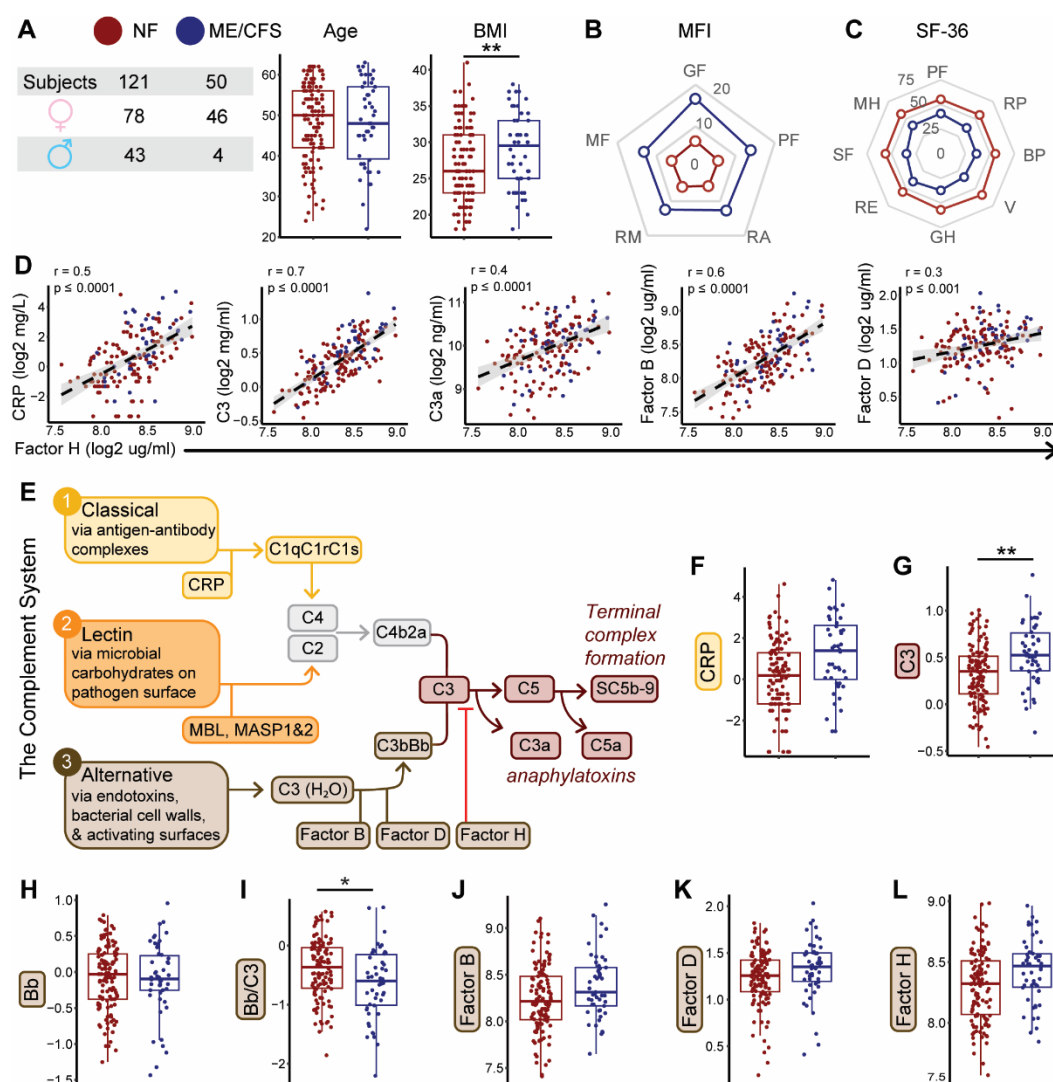
To extend our prior work, we investigate the association of plasma concentrations of complement proteins and related activation products with ME/CFS by identifying pQTLs among the 9146 SNPs that passed Affymetrix Human Immune and Inflammation Chip quality control [3]. We also identify correlations of complement proteins with measures of function, fatigue, and symptoms. Finally, we integrate these findings with the previously reported genetic variants associated with ME/CFS illness to identify a genetic subgroup of ME/CFS with complement dysregulation. In addition, we sought to validate our identified complement and disease associated pQTLs with the UK biobank's publicly available datasets for post-viral and fatigue-related phenotypes. This integrative approach has the potential to identify biologically distinct subgroups within the

heterogeneous ME/CFS population and should be helpful in studies of other chronic complex illnesses.

## 2. Results

### 2.1. Complement System Dynamics and Associations with Demographics and Disease Status

This study included 50 ME/CFS and 121 non-fatigued control Caucasian participants. While the two groups were comparable in age, significant differences were noted in sex distribution and body mass index (BMI) (Figure 1A). Among participants with ME/CFS, the median time since onset of fatigue was 8.97 years (range: 0.39 – 40.2 years), with 82.2% reporting a gradual illness onset [3]. Compared to the non-fatigued subjects, ME/CFS subjects had worse fatigue (higher scores) in all 5 domains of Multidimensional Fatigue Inventory (MFI-20, Figure 1B), greater symptom burdens in CDC symptom inventory (CDC-SI), and worse functional impairments (lower scores) in all 8 subscales of Short Form Health Survey (SF-36v2, Figure 1C).



**Figure 1.** Demographics and complement protein levels in ME/CFS compared to non-fatigued control subjects (A) Demographic comparisons of sex, age ( $p=0.96$ ), and BMI ( $p=0.007$ ) between non-fatigued controls (NF = red) and ME/CFS (blue) subjects ( $n = 171$  total; 50 ME/CFS, 121 NF). Statistical comparisons for age and BMI were performed using a Wilcoxon rank-sum test. (B) Radar plot showing mean scores on Multidimensional Fatigue Inventory (MFI-20) subscales between NF and ME/CFS subjects, where lower MFI-20 scores indicate better health scores.: GF = general fatigue, PF = physical fatigue, RA = reduced activity, RM = reduced motivation, MF = mental fatigue. (C) Radar plot showing mean T-scores on Short Form 36 survey (SF-36) subscales between NF

and ME/CFS subjects, where a T-score of 50 refers to the norm of the US general population and lower SF-36 T scores signify a worse functional score: PF = physical functioning, RP = role: physical, BP = bodily pain, V = vitality, GH = general health, RE = role: emotional, SF = social functioning, MH = mental health. **(D)** Pearson correlation dot plots depicting the extent of correlation of log<sub>2</sub>-transformed Factor H with (left to right): CRP ( $p=6.1E-12$ ), C3 ( $p=9.9E-28$ ), C3a ( $p=4.3E-9$ ), Factor B ( $p=1.4E-20$ ), and Factor D ( $p=6.9E-4$ ) in all subjects (NF = red, ME/CFS = blue). Linear regression lines with 95% confidence intervals are overlaid. **(E)** Schematic of the complement system and associated components, depicting the three activation pathways (color coded yellow – classical, orange – lectin, brown – alternative, and maroon – C3 and subsequent products). The same color coding is used in other figures to facilitate linkage back to the complement pathway. **(F-L)** Log<sub>2</sub>-transformed plasma levels of: **(F)** CRP (mg/L;  $p_{lin}=0.15$ ,  $p_{log}=0.14$ ), **(G)** C3 (mg/ml;  $p_{lin}=0.002$ ,  $p_{log}=0.002$ ), **(H)** Bb (mg/ml;  $p_{lin}=0.78$ ,  $p_{log}=0.77$ ), **(I)** Bb/C3 ( $p_{lin}=0.038$ ,  $p_{log}=0.038$ ), **(J)** Factor B (μg/ml;  $p_{lin}=0.71$ ,  $p_{log}=0.69$ ), **(K)** Factor D (μg/ml;  $p_{lin}=0.16$ ,  $p_{log}=0.14$ ), and **(L)** Factor H (μg/ml;  $p_{lin}=0.07$ ,  $p_{log}=0.07$ ) compared between NF and ME/CFS. Boxplots display the five-number summary: minimum, first quartile, median, third quartile, and maximum. The central rectangle spans from the first quartile to the third quartile (the interquartile range (IQR)), a segment inside the rectangle shows the median, the diamond shows the mean, the vertical lines (sometimes referred to as whiskers 1.5xIQR) are extended to the extrema of the distribution in the data set, and the outliers are values outside the whisker range. Statistical comparisons of circulating complement protein levels were performed using a covariate-adjusted (Table 1) linear ( $p_{lin}$ ) and logistic ( $p_{log}$ ) regression analyses. \* $p \leq 0.05$ , \*\* $p \leq 0.01$ .

Associations of the plasma concentrations of eight complement-related analytes spanning both core proteins and activation products of the alternative pathway – CRP, C3, Bb, C3a, C5a, Factor B, Factor D, Factor H, and SC5b-9 – with demographic covariates in all subjects, including age, sex, and BMI, are shown in Table 1 and Figure S1A-K. BMI was the most consistent covariate, showing moderate yet significant positive associations with 7 of the 9 measured proteins: CRP, C3, C3a, C5a, Factor B, Factor D, and Factor H (all  $p < 0.0001$ ), strongest ( $R^2$ ) with C3. In contrast Bb and SC5b-9 were not significantly associated with BMI, though the ratio of Bb to C3 was inversely associated. Sex was significantly associated with CRP and Factor B (Table 1, Figure S1I and J), with lower levels observed in males. Only Factor D levels were associated with age (Figure S1K). The demographic factors listed in Table 1 were considered as covariates in downstream analyses of the associated proteins.

**Table 1.** Association of demographic factors with plasma complement proteins and activation products (log<sub>2</sub> transformed).

Plasma component or fragment	Covariate	Coefficient	Std Error	Adjusted R <sup>2</sup>	F Statistic	p-value
CRP	BMI	0.169	0.024	0.230	47.47	1.3E-10
C3		0.033	0.004	0.269	63.11	2.7E-13
C3a		0.036	0.008	0.102	20.28	1.2E-05
C5a		0.025	0.005	0.106	21.06	8.7E-06
Factor B		0.026	0.005	0.142	28.32	3.3E-07
Factor D		0.020	0.004	0.113	22.62	4.2E-06
Factor H		0.023	0.004	0.181	38.47	4.2E-09
Bb/C3		-0.027	0.008	0.066	12.92	4.3E-04
CRP	Sex	0.645	0.318	0.019	4.10	4.5E-02
Factor B		0.159	0.061	0.034	6.84	9.8E-03
Factor D	Age	0.011	0.002	0.117	23.44	2.9E-06

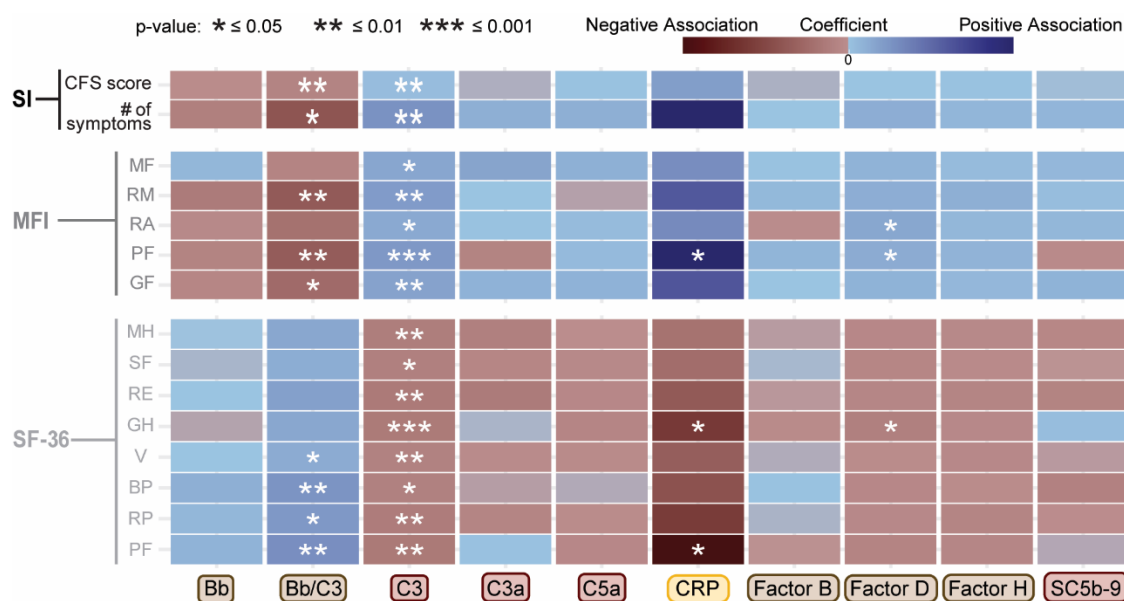
\*No association of Bb and Sc5b-9 with any of the demographic factors tested in this study.

We calculated correlations among all measured complement proteins across all subjects to evaluate protein interactions. Those with Factor H (Figure 1D) demonstrated the broadest pattern of correlation, including moderate relationships with CRP, C3, C3a, Factor B, and Factor D. The correlation of Factor H was much stronger with C3 and Factor B, and weaker with Bb ( $r = 0.11$ ,  $p = 0.12$ ). Given that the alternative complement cascade consists of multiple interdependent and tightly regulated components (Figure 1E), the majority of the pairwise comparisons showed significant positive associations.

To evaluate illness-associated differences in complement protein levels, we performed both linear and logistic regression analyses comparing protein concentrations between illness groups while adjusting for covariates. Linear regression assessed differences in protein concentrations as continuous outcomes, while logistic regression modeled the odds of ME/CFS as a function of protein levels. Box plots of log<sub>2</sub> transformed values stratified by disease status are shown in Figure 1F-L and Figure S1L-N. Despite several proteins having higher median values by at least 5% in ME/CFS compared to NF, only C3 and Bb/C3 ratio were significantly associated with ME/CFS in both models. The linear model indicates that C3 levels differ significantly between the two groups and the logistic model shows that higher C3 levels are associated with increased odds of ME/CFS.

## 2.2. Associations Between Complement System Components and Functional Health Scores

Results of the covariate-adjusted linear regression modeling between each analyte and the measures of function (SF-36), fatigue (MFI-20), and symptoms (CDC-SI) are shown as a heatmap Figure 2A and in correlation plots, Figure S2A-D. C3 shows the most consistent pattern among the tested complement proteins, with higher plasma levels associated with greater symptom burden, increased fatigue, and lower functioning scores (Figure S2B). CRP exhibited the largest beta coefficients across several MFI-20 and SF-36 domains, indicating that small increases in severity or functional impairment were associated with relatively large shifts in CRP levels; however, the corresponding  $R^2$  values remained low (Figure S2C;  $R^2 \sim 0.32$ ), suggesting that despite strong directional effects, the symptom scores explained only a modest proportion of the overall variance in participants' CRP levels. Factor D and the Bb-to-C3 ratio also showed significant associations with some MFI and SF-36 metrics. In contrast, Bb, C3a, C5a, Factor B, Factor H, and SC5b-9 showed no significant associations with any individual scores.



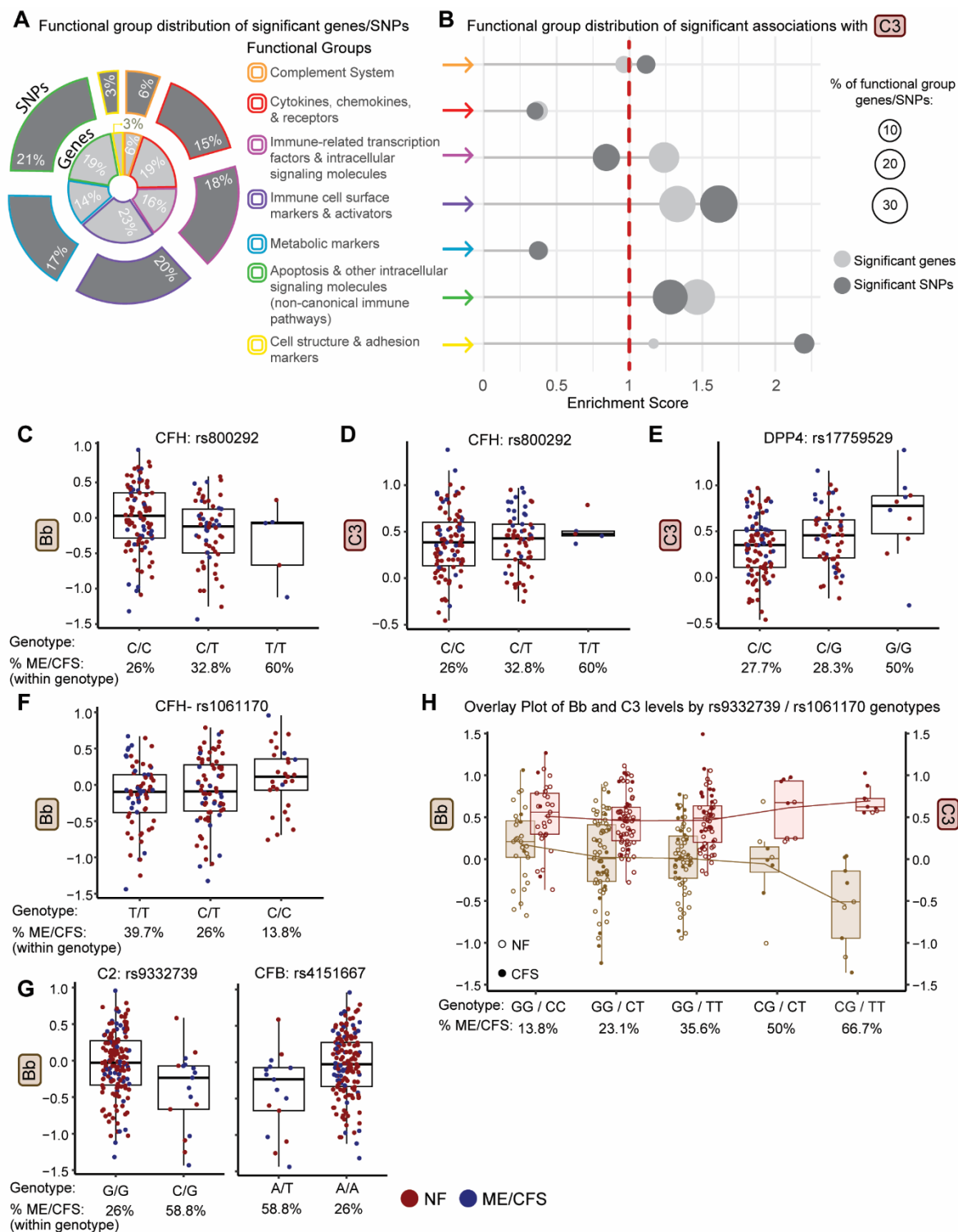
**Figure 2.** Association of circulating complement protein levels with participants' function and symptoms scores. Heatmap depicting covariate-adjusted associations between plasma complement protein levels and scores from three questionnaires: Symptom Inventory (CDC-SI, top), Multidimensional Fatigue Inventory (MFI-20, middle),

and Short Form 36 survey (SF-36 T scores, bottom). Lower SF-36 T scores signify a worse functional score, while lower MFI-20 and CDC-SI scores indicate better health scores. Rows represent individual domains or subscales for each instrument, and columns represent complement-related proteins. Color represents the coefficient, or magnitude, of each association, with blue indicating a positive association, while red/maroon indicates a negative association. Statistical comparisons were performed using covariate-adjusted linear regression analysis. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . (SI score= CDC Symptom Inventory; MFI: MF = mental fatigue, RM = reduced motivation, RA = reduced activity, PF = physical fatigue, GF = general fatigue; SF-36: MH = mental health, SF = social functioning, RE = role: emotional, GH = general health, V = vitality, BP = bodily pain, RP = role: physical, PF = physical functioning). NOTE: Complement protein labels are color coded to match components of complement system illustrated in Figure 1E.

### 2.3. Identification of Genetic Variants Impacting Plasma Levels of Complement Proteins

To investigate if genetic variants influence circulating complement protein levels, we performed a quantitative trait locus (pQTL) analysis using all 9,146 SNPs previously analyzed in this population [3], identifying 3,192 SNPs significantly associated with at least one complement protein ( $p < 0.05$ ). Data on the SNPs associated with each protein is presented in the supplement (Table S1-10). To distill meaningful biologic themes from this large data set, the 776 SNPs with  $p < 0.01$ , representing 359 genes, were categorized into one of seven curated functional groups [(1) direct association with the complement system (20 genes), (2) cytokines, chemokines, and their receptors (69 genes), (3) immune-associated transcription factors and intracellular immunomodulators (57 genes), (4) immune cell surface markers and activators (82 genes), (5) metabolic markers (50 genes), (6) apoptosis and other intracellular signaling molecules involved in non-canonical immune pathways (70 genes), and (7) cell structure and adhesion markers (11 genes); Table S11]. The distribution of the 776 significant SNP and 359 genes by functional group is shown graphically in Figure 3A.

This framework enabled analysis of functional group enrichment scores for each complement protein association (C3 in Figure 3B, and all other proteins in Figure S3A-I). For C3, we observed over-representation in the functional groups immune cell surface markers and activators, apoptosis, and cell structure and adhesion. Statistical parameters for each protein's most significant associations are found in Supplemental Tables S1-S10. Two SNPs, rs800292 in *CFH* and rs17759529 in *DPP4*, linked to Bb and C3 levels are among the most significant in this pQTL screen. Genotypes of rs800292/*CFH*, previously implicated in complement regulation, showed a stepwise decrease in Bb levels with the T allele (Figure 3C). Interestingly, rs800292 shows a similar decrease in Bb/C3 ratios (Figure S4A), with an inverse trend seen in genotype-stratified C3 levels (Figure 3D). The G allele of rs17759529 in *DPP4*, a top-ranked pQTL for CRP and C3 in the over-represented immune cell surface markers functional group, had higher C3 levels in homozygous carriers (Figure 3E). Notably, *DPP4* is a multifunctional protease expressed on immune cells and endothelium that has been implicated in SARS-CoV-2 [33]. Additionally, the rs17611 SNP in *C5* was among our top-ranked pQTLs for C5a in this population (Figure S4B) in agreement with previous reports in independent cohorts [34,35]. The major allele T in rs1061170 (*CFH*) was associated with lower Bb levels (Figure 3F). Bb levels were similarly significantly lower in individuals carrying minor C allele rs9332739 (*C2*) and T allele rs4151667 (*CFB*), two SNPs in linkage disequilibrium (LD) (Figure 3G). The box plots in Figures 3C-3G also show the skewed distribution of ME/CFS cases by genotype of these complement pQTLs, suggesting an interaction between complement pQTLs and ME/CFS risk.



**Figure 3.** Functional group distribution of complement protein-associated SNPs/genes and complement protein levels by genotype and illness status. **(A)** Donut plot summarizing the distribution across 7 functional groups of single-nucleotide polymorphisms (SNPs - dark grey outer ring) and corresponding genes (light grey inner ring) associated with complement proteins ( $p \leq 0.01$  after covariate-adjustment). Functional group assignments were based on annotated gene functions for gene and associated SNPs (Tables S1-11). **(B)** Lollipop plot showing functional group enrichment among SNPs and genes significantly associated with C3 plasma levels. Dot size reflects the proportion of significant functional group genes (light grey) and SNPs (dark grey). Enrichment score  $> 1$  (dotted red line) indicates overrepresentation. **(C-G)** Boxplots of plasma protein levels ( $\log_2$ , mg/ml) by genotype. Dot color indicates disease status (NF = red; ME/CFS = blue); x-axis = percentage of ME/CFS subjects within each genotype group. **(C)** Bb for rs800292/CFH genotypes ( $p=0.003$ ); **(D)** C3 for rs800292/CFH genotypes

( $p = 0.61$ ); (E) C3 for rs17759529/*DPP4* genotypes ( $p=0.0002$ ); (F) = Bb for rs1061170/*CFH* genotypes ( $p=0.025$ ); (G) left: Bb for rs4151667/*CFB* genotypes ( $p=0.002$ ), right: Bb for rs9332739/*C2* genotypes ( $p=0.002$ ). (H) Overlay boxplots of Bb (brown, primary y-axis, log<sub>2</sub>, mg/ml) and C3 (maroon, secondary y-axis, log<sub>2</sub>, mg/ml) by genotype combinations of rs9332739/*C2* and rs1061170/*CFH* (NF = empty dot, ME/CFS = filled dot). Boxplots represent the median  $\pm$  25<sup>th</sup> and 75<sup>th</sup> quartiles. Whiskers represent 1.5x the interquartile ranges. Outliers are values outside the whisker range.

#### 2.4. Identification of Genetic Variants Impacting Both Plasma Levels of Complement Proteins and Disease Status

We identified 16 SNPs found in both the 3,192 protein-associated SNPs and the 48 previously reported ME/CFS-associated SNPs [3], 11 of which act in the complement cascade (Table 2). The additional immune and inflammatory pathways represented indicate genetic risk for ME/CFS beyond the complement system.

Among the 16 SNP associations, four were predicted to act in cis (defined as variants located within 1Mb of the transcription start site of the impacted protein's associated gene), such as variants located in *C2*, *CFB*, or *CFH* impacting levels of Bb, Factor B, or Factor H, respectively. The remaining 12 variants were predicted to be trans-pQTLs, acting elsewhere in the genome (e.g. *SERPINA5* impacting C3, *GRK4* impacting C3, and *PDE4D* impacting C3a and C5a). Of note, rs3020729 (*CD8A*, miRNA binding) was significantly associated with higher levels of CRP and C3 levels, and rs2277680 (*CXCL16*, missense) impacted Factor H and Factor B levels (Table 2). By contrast, rs9550987 (*TNFRSF19*, missense) was associated with decreased levels of C3 and Factor H (Table 2).

**Table 2.** Genetic variants impacting both disease status and plasma levels of measured complement proteins and activation factors.

Pathway	Chromosome #: Gene name	SNP rsID	ME/CFS risk allele (nt)	Variant consequence	Impacted complement protein (cis/trans)	$\beta \pm SE$	p-value
Complement cascade	6: <i>C2</i>	rs9332739~	Minor (C)	Missense (E318D)	Bb/C3	-0.49 $\pm$ 0.13	1.9E-04
					Bb (cis)	-0.36 $\pm$ 0.12	0.002
					Factor B (cis)	-0.23 $\pm$ 0.08	0.007
	6: <i>CFB</i>	rs4151667~	Minor (T)	Missense (L9H)	Bb/C3	-0.49 $\pm$ 0.13	1.9E-04
					Bb (cis)	-0.36 $\pm$ 0.12	0.002
					Factor B (cis)	-0.23 $\pm$ 0.08	0.007
		rs641153	Major (C)	Missense (R32L)	Factor B (cis)	0.24 $\pm$ 0.07	0.001
					Bb/C3	-0.24 $\pm$ 0.10	0.021
					Factor D (trans)	-0.11 $\pm$ 0.05	0.046
	1: <i>CFH</i>	rs7529589*	Major (G)	Intronic	Factor H (cis)	0.06 $\pm$ 0.03	0.020
					Bb (trans)	0.10 $\pm$ 0.05	0.036
		rs1061147*	Major (G)	Codon-synonymous (A307A); splicing regulation	Bb (trans)	0.12 $\pm$ 0.05	0.015
					Bb (trans)	-0.19 $\pm$ 0.06	0.003
		rs800292	Minor (T)	Missense (V62I)	Bb/C3	-0.21 $\pm$ 0.07	0.003
					Bb (trans)	0.11 $\pm$ 0.05	0.025
		rs1061170*	Major (T)	Missense (H402Y)	Bb (trans)	0.12 $\pm$ 0.05	0.014
Bb (trans)	0.12 $\pm$ 0.05				0.014		
3: <i>MASP1</i>	rs3774268	Minor (T)	Missense (S445R)	Bb/C3	-0.24 $\pm$ 0.08	0.003	
				Bb (trans)	-0.17 $\pm$ 0.07	0.026	
14: <i>SERPINA5</i>	rs6115	Minor (G)	Missense (S64N)	Bb/C3	-0.11 $\pm$ 0.05	0.047	

		<u>rs6108</u>	Minor (A)	UTR-3; miRNA binding	C3 (trans)	0.07 ± 0.03	0.026
T cell signaling	2: <i>CD8A</i>	<u>rs3020729</u>	Major (T)	UTR-3; miRNA binding	CRP (trans)	-0.67 ± 0.25	0.008
					C3 (trans)	-0.10 ± 0.05	0.028
Chemokine	17: <i>CXCL16</i>	rs2277680	Major (G)	Missense (I142T; A200V)	Factor H (trans)	-0.06 ± 0.03	0.024
					Factor B (trans)	-0.08 ± 0.04	0.031
G-protein coupled receptor signaling	5: <i>PDE4D</i>	<u>rs2014012</u>	Major (A)	Intronic	C3a (trans)	-0.13 ± 0.06	0.036
					C5a (trans)	-0.09 ± 0.04	0.038
	4: <i>GRK4</i>	<u>rs1801058</u>	Major (C)	Missense (V486A); splicing regulation	C3 (trans)	-0.07 ± 0.03	0.034
TNF super family signaling	13: <i>TNFRSF19</i>	<u>rs9550987</u>	Minor (T)	Missense (S31T); splicing regulation	C3 (trans)	0.09 ± 0.03	0.011
					Factor H (trans)	0.07 ± 0.03	0.027

SNP IDs underlined indicate pQTLs that have not been previously reported (no match in the QTL database: <http://www.mulinlab.org/qtlbase>).~ and \* indicate SNPs that are in LD

As shown in Figure 3F, the ME/CFS risk-associated major allele T in rs1061170 (*CFH*) was associated with lower Bb levels. Bb levels were similarly lower in individuals carrying the ME/CFS risk alleles at cis-acting rs9332739 (*C2*) and rs4151667 (*CFB*), two SNPs in linkage disequilibrium (LD) (Figure 3G). The inverse trends in genotype-stratified C3 levels and Bb/C3 ratio values for rs800292 (*CFH*) shown in Figure 3C, are also seen with the rs9332739 (*C2*) risk allele (Figure S4C and D). The five genotype combinations for these *C2* and *CFH* variants (Figure 3H) show decreasing Bb and increasing C3 levels with increasing risk alleles, with the double risk genotype (GC-TT) exhibiting the lowest Bb and highest C3 levels as well as the highest percentage of ME/CFS in the genotype combination. The correlation analysis of disease-associated SNP beta values and disease odds ratios (Figure S4E and F) agrees with this pattern; variants increasing C3 expression aligned with increased ME/CFS risk (Figure S4E,  $R^2 = 0.78$ ), while those associated with higher Bb showed inverse trends (Figure S4F,  $R^2 = -0.26$ ).

### 2.5. Use Of Genotype-Defined Complement pQTLs to Identify ME/CFS Subgroups

As the levels of complement proteins are associated with ME/CFS, we assessed their performance as predictive markers using an ROC analysis of all 171 subjects (50 ME/CFS, 121 NF controls, Figure S5A). CRP, Factor B, and C3 had the highest predictive power, with AUCs of 0.75, 0.75, and 0.69, respectively (Figure S5A); only C3 was statistically significant. To incorporate the impact of complement-associated pQTLs, we conducted ROC analyses for complement proteins with ME/CFS as the outcome, stratified by genotype subgroups for single SNPs, selected based on significant associations with complement proteins or disease status (Table 2 and Table S1-10) and restricted to those within the complement system functional group. Results were filtered for AUC >0.75,  $p < 0.05$ , and sufficient group size ( $n = 15$ ) and are summarized in Table 3. Heterozygotes for two SNPs, rs9332739 in *C2* and rs800292 in *CFH*, previously examined in Figure 3, showed a high AUC (Table 3). This SNP genotype combination was selected for subgroup analysis.

**Table 3.** Summary of ROC analysis stratified by genotype-defined pQTLs for identifying ME/CFS subgroups with predictive complement protein markers.

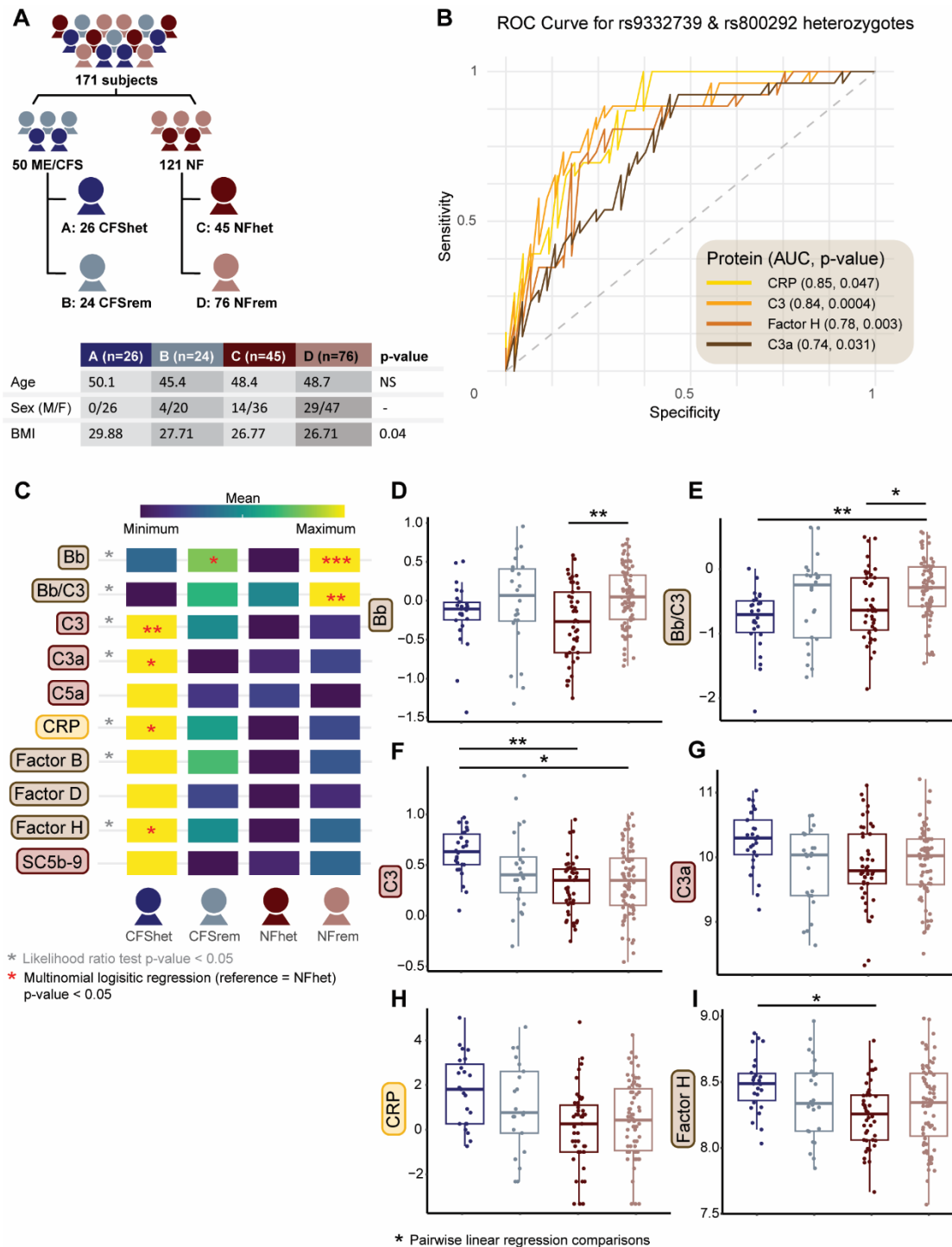
Variant	Gene	Genotype	Test variable	Covariates	# of CFS	# of NF	# of subjects	AUC ± SE	95% CI	P-value
<b>Single marker analysis</b>										

rs933273 9	C2/CF B	CG	Factor H	BMI	10	7	17	0.89 ± 0.09	0.71 - 1	0.05
rs800292	CFH	CT	C3	BMI	20	41	61	0.82 ± 0.06	0.74 - 0.94	0.001
			Factor H					0.78 ± 0.06	0.7 - 0.94	0.007
			CRP	Sex, BMI	17	40	57	0.84 ± 0.05	0.72 - 1	0.04
rs106117 0		TT	Factor B	Sex, BMI	27	40	67	0.81 ± 0.06	0.7 - 0.92	0.03
rs108015 55		GG	Factor B	Sex, BMI	26	38	64	0.81 ± 0.06	0.7 - 0.92	0.03
rs395544		CC	Factor B	Sex, BMI	20	36	56	0.83 ± 0.05	0.73 - 0.94	0.03
rs106548 9	GG	C3	BMI	27	61	88	0.75 ± 0.06	0.64 - 0.86	0.02	
rs713597 5	C1R	AG	C3	BMI	16	43	59	0.76 ± 0.08	0.6 - 0.92	0.01
rs725706 2	C3	CC	C3	BMI	20	43	63	0.78 ± 0.06	0.65 - 0.91	0.006
rs224139 3		CC	C3	BMI	18	38	56	0.81 ± 0.06	0.69 - 0.93	0.003
rs224139 2		CC	C3	BMI	19	43	62	0.78 ± 0.07	0.65 - 0.91	0.006
rs703767 3	C5	CT	C3	BMI	16	45	61	0.82 ± 0.05	0.72 - 0.93	0.005
			Bb/C3					0.81 ± 0.06	0.69 - 0.93	0.006
rs261753	C9	CC	CRP	Sex, BMI	31	76	107	0.78 ± 0.05	0.69 - 0.87	0.04
rs198615 8	CR1	CT	C3	BMI	11	23	34	0.83 ± 0.09	0.65 - 1	0.009
<b>2-marker analysis</b>										
rs933273 9 & rs800292	C2/CF H	CG/CT	CRP	Sex, BMI	23	43	66	0.85 ± 0.05	0.76 - 0.94	0.046
			C3	BMI	26	45	71	0.84 ± 0.05	0.74 - 0.94	0.0004
			Factor H					0.78 ± 0.06	0.67 - 0.89	0.003

For single marker and 2-marker analyses, test variables with AUC >0.75 and p < 0.05 are listed.

Subjects were stratified into four groups based on heterozygosity at rs9332739 in C2 and rs800292 in CFH. ME/CFS participants with heterozygous genotypes at one or both SNPs were assigned to CFShet (n = 26, Group A), while the remaining ME/CFS cases were designated as CFSrem (n = 24, Group B). Non-fatigued controls were similarly divided into NFhets (n = 45, Group C) and NFrem (n = 76, Group D), with the distribution and subgroup demographics shown in Figure 4A. Because subgrouping was defined by heterozygosity, two individuals with the double risk genotype for rs800292 but no risk allele at rs9332739 were assigned to the CFSrem and NFrem groups. The CFShet group contained no male participants, and BMI was significantly different between the CFShet group and the other three groups. CRP and C3 exhibited improved AUCs in the heterozygous-defined CFShets vs. NFhets comparison, with AUC values of 0.85 and 0.84, respectively (Table 3, Figure 4B),

while Factor H and C3a also had enhanced predictive power (0.78 and 0.74 AUC, respectively) over the non-stratified model (Figure S5A).

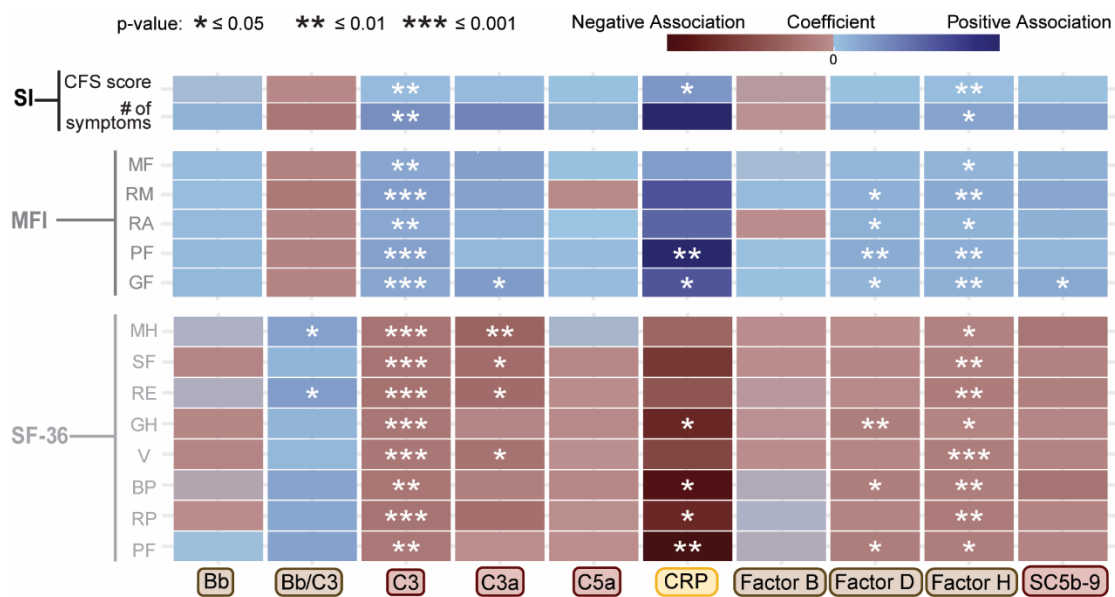


**Figure 4.** Stratification of ME/CFS and non-fatigued control subjects using complement protein-associated pQTL genotypes. **(A)** Subgroup distribution and demographics (age, sex, BMI) based on genotype status of rs9332739 (C2) and rs800292 (CFH). Group A (CFShet, dark blue) = ME/CFS heterozygous at one or both loci; Group B (CFSrem, light blue) = Remainder of ME/CFS subjects; Group C (NFhet, dark red) = NF controls heterozygous at one or both loci; Group D (NFrem, pink) = remainder of NF controls. **(B)** Complement protein ROC curves for ME/CFS, restricted to heterozygous groups (Group A (CFS het) and Group C (NFhet)). **(C)** Heatmap based on the log<sub>2</sub>-transformed mean range for each complement protein, low (blue) to high (yellow). Mean values for genotype-defined subgroups (CFShet, CFSrem, NFhet, NFrem) are shown. Asterisks indicate statistical

significance, adjusted for covariates (grey = likelihood ratio test with  $p \leq 0.05$ , red = subgroup-specific multinomial logistic regression with reference = NFhet and  $p \leq 0.05$ ). **(D-I)** Boxplots showing distributions for plasma levels of complement protein (log<sub>2</sub>, mg/ml) by genotype group, CFShet = dark blue, CFSrem = light blue, NFhet = dark red, NFrem = pink. **(D)** Bb (NFhet vs. NFrem  $p=0.002$ ), **(E)** Bb/ C3 (CFShet vs. NFrem  $p=0.003$ ; NFhet vs. NFrem  $p=0.03$ ), **(F)** C3 (CFShet vs. NFhet  $p=0.005$ , CFShet vs. NFrem  $p=0.02$ ), **(G)** C3a, **(H)** CRP, and **(I)** Factor H (CFShet vs. NFhet  $p=0.05$ ). Boxplots display the five-number summary: minimum, first quartile, median, third quartile, and maximum. The central rectangle spans from the first quartile to the third quartile (the interquartile range (IQR)), a segment inside the rectangle shows the median, the diamond shows the mean, the vertical lines (sometimes referred to as whiskers  $1.5 \times \text{IQR}$ ) are extended to the extrema of the distribution in the data set, and the outliers are values outside the whisker range. Statistical comparisons of circulating complement protein levels (black asterisk) were performed using a covariate-adjusted (Table 1) pairwise linear regression analysis. \* $p \leq 0.05$ , \*\* $p \leq 0.01$ , \*\*\* $p \leq 0.001$ .

To assess complement protein differences across genotype-stratified illness groups, we applied three statistical approaches: a likelihood ratio test to evaluate overall model fit, a multinomial logistic regression to identify proteins associated with subgroup assignment (using NFhet as the reference), and pairwise covariate adjusted linear regressions to examine direct groupwise differences. The likelihood ratio test identified seven proteins – Bb, Bb/C3, C3, C3a, CRP, Factor B, and Factor H – that significantly improved the model, indicating that these proteins contribute meaningfully to group classification (Figure 4C). Multinomial regression revealed that C3, C3a, CRP, and Factor H levels were significantly associated with increased odds of belonging to the CFShet group relative to NFhet (Figure 4C), suggesting potential utility as subgroup classifiers. Pairwise linear regression comparisons confirmed several between-group differences (Figure 4D-I), including lower Bb and Bb/C3 in NFhet compared to NFrem, and lower Bb/C3 in CFShet compared to NFrem. The CFShet group also demonstrated higher C3 and Factor H levels relative to NFhet, with CFShet C3 levels significantly higher than NFrem as well. Collectively, these complementary analyses support a robust association between circulating complement proteins and genotype-defined subgroups – even after adjusting for covariates – highlighting their potential role in distinguishing biologically meaningful ME/CFS subgroups.

We examined the covariate-adjusted association between complement protein levels and scores of symptoms and function for the heterozygous subgroups (CFShet and NFhet, Figure 5A, Figure S5B-D). The covariate-adjusted regression models in this reduced population ( $n = 71$ ) revealed multiple significant associations between complement protein levels and scores in CDC-SI, MFI-20, and SF-36. In the total study sample population (Figure 2), only four complement-related measurements (Bb/C3, C3, CRP, and Factor D) had significant associations compared to the seven complement proteins with significant associations across at least one of the three survey domains in the genotype-restricted groups (Figure 5A). These included newly significant relationships for C3a, Factor H, and SC5b-9 (Figure 5A, Figure S5B-D). For the total and genotype-restricted samples, the correlations followed the same direction. While exploratory, this expanded set of associations suggest that complement dysregulation may be more tightly linked to reported functional impairments in genetically defined subgroups, reinforcing the relevance of host genotypes in shaping disease heterogeneity in ME/CFS.



**Figure 5.** Associations between complement protein levels and symptom/function scores for genotype-restricted subgroup (CFShet, NFhet, n = 71). Heatmap of covariate-adjusted associations between complement protein levels and scores: Symptom Inventory (SI, top), Multidimensional Fatigue Inventory (MFI, middle), and Short Form 36 survey (SF-36 T scores, bottom). Rows represent individual domains or subscales for each survey, and columns represent complement-related proteins. Color represents the coefficient, or magnitude, of each association, with blue indicating a positive association, while red indicates a negative association. Statistical comparisons were performed using covariate-adjusted linear regression analysis. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . (SI = CDC Symptom Inventory; MFI: MF = mental fatigue, RM = reduced motivation, RA = reduced activity, PF = physical fatigue, GF = general fatigue; SF-36: MH = mental health, SF = social functioning, RE = role: emotional, GH = general health, V = vitality, BP = bodily pain, RP = role: physical, PF = physical functioning).

## 2.6. Comparison of Significant SNPs Between Our Population and Fatigue-Related Phenotypes in the UK Biobank

To evaluate the consistency and broader applicability of a pQTL-based approach for identifying genetic contributors to ME/CFS, we compared SNPs associated with each circulating complement protein (Table S1-10, top 50 SNPs from each table with the lowest p-values) identified in our population to SNPs associated with fatigue traits in the publicly available UK biobank GWAS data for chronic fatigue syndrome, post viral fatigue syndrome, a broader fatigue and malaise population, and Ever CFS (individuals reporting recovery from ME/CFS). Six SNPs across 5 genes overlapped between our pQTL dataset and the UK biobank fatigue-related traits (Table 4), mapping to complement-related genes *CFB*, *CFH*, and *C2*, as well as *MMP10* (involved in the breakdown of extracellular matrix) and *PIK3R5* (which plays roles in cell growth, proliferation, differentiation, motility, survival, and intracellular trafficking). Four of the 6 SNPs (3 genes) were among those we identified as both pQTLs and ME/CFS-associated (rs641153, rs800292, rs9332739 and rs4151667). Interestingly, rs394811 in *PIK3R5* was the only SNP to overlap with the UK biobank's designated CFS group. In contrast, SNPs in *CFB*, *CFH*, and *C2* appeared in the broader fatigue and malaise and post viral fatigue phenotypes. Moreover, the direction of effect for each overlapping SNP – including risk allele identity and odds ratio – was consistent between the UK biobank phenotypes and our population, providing some replication of the biological relevance of these shared associations.

**Table 4.** Overlapping SNP associations between circulating complement proteins and UK biobank fatigue-related phenotypes.

UK Biobank fatigue-related phenotype	Chr #: Gene name	Variant	Variant consequence (reference nt - variant nt)	UK biobank: cases / healthy controls	SNP - UK biobank phenotype odds ratio	SNP - UK biobank phenotype odds ratio confidence interval	SNP - UK biobank phenotype p-value	SNP - Impacted Complement protein p-value
Chronic Fatigue Syndrome	17: <i>PIK3R5</i>	rs394811	Synonymous (G-A)	2047 / 307792	0.847	0.75 - 0.96	0.007	0.002 (C3)
Post Viral Fatigue Syndrome	6: <i>CFB</i>	rs641153	Missense (G-A)	4363 / 216118	0.571	0.38 - 0.86	0.005	0.001 (CFB)
	11: <i>MMP10</i>	rs17293607	Missense (C-T)	4360 / 216027	0.920	0.86 - 0.98	0.009	0.004 (C5a)
41202/R53 Fatigue & Malaise	1: <i>CFH</i>	rs800292	Missense (G-A)	2132 / 283735	0.754	0.61 - 0.94	0.010	0.003 (Bb)
Ever CFS	6: <i>C2</i>	rs9332739	Missense (G-C)	2547 / 121864	1.186	1.05 - 1.34	0.008	0.002 (Bb)
	6: <i>CFB</i>	rs4151667	Missense (T-A)	2547 / 121863	1.185	1.05 - 1.34	0.009	0.002 (Bb)

### 3. Discussion

This study devised a novel pQTL-based approach to address clinical heterogeneity in ME/CFS using genotype and pathway-focused stratification. Our analysis focused on complement pQTLs and yielded several noteworthy findings. First, we observed significantly higher levels of C3 in ME/CFS compared to NF controls, alongside a decrease in the ratio of Bb to C3, suggesting altered alternative pathway activation in this disease. Second, we identified multiple complement pQTLs that were also associated with disease status. These pQTLs included polymorphisms in *C2/CFB* (rs9332739 and rs4151667 in LD) and *CFH* (rs800292). The ability of complement protein levels to distinguish between ME/CFS and NF controls was improved when limiting the analysis to the genotype-restricted subgroup. Cases and controls in this subgroup not only differed in C3 and Factor H levels but also showed distinct patterns across multiple analyses that improved the model's ability to distinguish between the four subgroups and improved the odds of subgroup assignment, with proteins in the genotype-restricted group also exhibiting more extensive associations with functional impairment and patient health scores. Additionally, several pQTLs that we identified in our population with functional consequences (non-synonymous SNPs: rs4151667, rs9332739, rs641153, rs800292, rs394811 and rs17293607) were found to overlap with SNPs that were significantly associated with fatigue-related phenotypes listed in the UK biobank, demonstrating how integrating approaches may help resolve inconsistencies across genetic studies of this heterogeneous disease. These findings stress a link between complement system genetics, complement protein dysregulation, and ME/CFS pathophysiology, while also providing an approach for the much-needed subgrouping of heterogeneous diseases such as ME/CFS.

Our association results reinforce the need to adjust data for covariates, such as age, sex, or BMI when evaluating differences in complement components between ME/CFS and healthy controls; adjustments likely important for other protein associations as well. Our findings of the influence of age, sex, and BMI on the level of these proteins are largely in agreement with recently published

results [36–38], requiring them to be used as covariates in the analysis of complement proteins using genetic or non-genetic models. The complement dysregulation identified in our study of ME/CFS persisted after covariate adjustment, highlighting its disease-specific nature rather than demographic artifact. While age is known to influence certain classical complement components, such as declining C1q levels in older adults [39,40], we only included age as a covariate in models where it was empirically significant for our specific population, such as for Factor D. In contrast, BMI emerged as a consistent covariate across multiple complement proteins – most notably C3 – likely reflecting adipose-driven complement protein production [41–44]. This is particularly relevant for ME/CFS, where symptom severity and metabolic shifts may influence body composition. Of note, earlier studies that did not control for BMI reported no differences in C3 between ME/CFS and controls [20,45]. Sex was also considered when statistically warranted (Factor B and CRP) and adjustment ensured complement differences were not driven by the unbalanced sex ratio observed in ME/CFS [46]. Sex hormones and x-linked immune genes shape distinct immunologic profiles [47], and women generally exhibit lower baseline complement levels [48,49]. Notably, the CFS<sub>het</sub> subgroup contained no male participants, suggesting potential sex-dependent genotype effects that warrant future stratified analyses, in line with other recent suggestions of ME/CFS immune subtyping by sex [50,51].

C3 is the central protein in the complement cascade regulated by three major pathways (classical, lectin and alternate pathways) leading to a pathway-dependent C3 convertase converting C3 to C3b and C3a. The classical/lectin pathways generate C3 convertase by combining C4bC2b, whereas the alternative pathway contributes to C3 convertase (C3bBb) by binding of C3b to Bb (Bb results from cleavage of Factor B by Factor D). An increase in plasma Bb is considered as a marker of complement activation, whereas a decrease in Bb indicates complement inhibition resulting from the inhibitory functions of Factor H [18]. High C3 and C3a, along with high CRP (previously reported in other ME/CFS populations [52,53]) hint at ongoing innate immune activation; however, concurrent high Factor H (an alternative pathway inhibitor) and low Bb suggest increased complement inhibition in our ME/CFS subjects. This could reflect a compensatory response or genetically driven dampening of the alternative pathway in ME/CFS, especially in variant-carrying subgroups. Of note, a recent plasma proteomic analysis in ME/CFS suggests a similar hypotheses of attenuated complement activity, reflecting a compensatory mechanism and subsequent exhausted immune state in ME/CFS [54]. Integrating genetic variant data, we found subjects carrying the risk allele in *C2/CFB* and *CFH* (rs9332739 and rs1061170) had lower levels of Bb in plasma, and that Bb levels could be impacted by the pQTL rs800292 in *CFH*, although this latter trans-pQTL was not directly associated with illness (Armitage  $P=0.09$ ). Notwithstanding, heterozygotes for rs9332739 in *C2* and rs800292 in *CFH* resulted in the highest AUC for prediction of illness status. Both rs9332739 (E318D) in *C2* and rs800292 (V62I) in *CFH* are missense pQTLs for Bb and both are reported to be protective for AMD [27–32,55].

There is limited information on the functional role of these pQTLs in the alternate pathway. Other studies have documented their association with diseases that have seemingly unrelated clinical presentations and pathologies. For example, studies of AMD and atypical hemolytic uremic syndrome (aHUS), clinically different illnesses but both having overactivity of the alternative pathway, identified defective regulation of the alternative pathway [56,57]. This differs from our finding that at least a subset of ME/CFS subjects show inhibition of the alternative pathway through what appears to be a genetically determined low level of Bb. Moreover, when *C2* and *CFH* variants were considered together, Bb and C3 levels showed stepwise opposing trends across the five genotype combinations, most stark in the genotype group with the highest ME/CFS proportion. These reciprocal trends point to a genetically encoded imbalance between complement precursor (C3) and activation product (Bb), potentially reflecting dysregulated convertase activity in ME/CFS. Correlation analyses between C3 or Bb protein levels and disease odds ratios revealed consistent directional concordance across diverse genetic associations, reinforcing a model in which ME/CFS genetic risk is linked to elevated upstream complement activity skewed toward C3 and away from Bb.

Higher complement inhibition and thus reduced Bb, as we see in this study, is also in agreement with reported inhibition of the alternate pathway by the minor allele of rs800292 in *CFH*, due to the strong binding of the variant to C3b [58,59]. Factor H and C3 protein were highly correlated, perhaps resulting from increased inhibition of alternate pathway by Factor H [60], and both were present in high levels in ME/CFS. Understanding these biological relationships is essential for interpreting circulating levels of complement proteins in ME/CFS, as disease-associated changes may reflect both direct dysregulation and broader shifts in pathway dynamics. For example, rs800292 in *CFH* is reported as a potent regulator of matrix metalloproteinase-8 (MMP-8), a proinflammatory enzyme conferring risk for cardiovascular diseases, and carriers of the rs800292 minor allele were found to have decreased levels of serum MMP-8 [61]. *CFH* belongs to a family of 7 genes, denoted as Regulators of Complement Activation (RCA), spanning a region of 360 kbp in chromosome 1 [59]. Mutations seen in RCA can act as both activators and inhibitors, offering the opportunity to regulate their role in different diseases [62]. Notably, we identified one cis-acting pQTL for Factor H, negatively regulating rs1065489 in *CFH*. This study also identified four disease-associated pQTLs for C3 (rs6108 in *SERPINA5*, rs3020729 in *CD8A*, rs1801058 in *GRK4*, rs9550987 in *TNFRSF19*). Although C3 levels were significantly different between illness groups, we did not identify any co-localized cis-QTL impacting both illness status and C3 level; however, circulating C3 levels were decreased by the minor alleles of three cis acting intronic pQTLs (rs7257062, rs2241393, rs2241392), all of which performed as strong predictors of illness status, with AUC > 0.78 and p-values below 0.006. Subjects homozygous for major alleles of these three variants had higher levels of C3, but for unclear reasons were not associated with ME/CFS. Of particular interest among pQTLs not associated with ME/CFS are several pQTLs (rs17611, rs7037673, rs1468673, rs992670, rs7040319, rs4837805, rs17220750, rs2416810, rs2300939, rs7026551) in the *TRAF1/C5* region in chromosome 9 that impact circulating C5a levels. Interestingly, QTLbase home database (www.mulinlab.org) lists rs7037673 as a pQTL for C5 expression in the prefrontal cortex, supporting further investigation into complement-mediated neuroinflammation in ME/CFS pathogenesis [63]. This variant also lies near *TRAF1*, a key component of the TNF signaling cascade that promotes inflammatory cytokine expression, suggesting a potential link between complement activity and neuroimmune signaling in ME/CFS. Overall, this approach highlights the potential of complement-associated pQTLs to delineate biologically meaningful markers and targets within ME/CFS.

Dysregulated complement activity in ME/CFS mirrors persistent complement activation, impaired metabolic and immune responses, and altered inflammatory cytokine profiles, consistent with chronic inflammation and incomplete post-infection resolution illnesses [64–73] reported in other infection-associated chronic conditions and illnesses (IACCIs). Understanding the nuanced differences in complement profiles across IACCIs may help explain why some individuals recover while others develop chronic sequelae. Related but mechanistically distinct complement system patterns are observed in other chronic immune or neurological conditions: in lupus, low CRP with C3/C4 consumption impairs nuclear antigen clearance, whereas in ME/CFS, elevated CRP alongside high Factor H and reduced Bb suggests suppressed clearance and unresolved inflammation driven by host genetics [74–76]. In multiple sclerosis (MS), complement activation contributes to demyelination, with high Factor H observed in some patients, despite a clear complement locus yet to be identified—consistent with predominantly environmental/secondary involvement [79,80]. In Alzheimer's disease, complement-associated variants and insufficient control of complement-mediated synaptic pruning implicate complement in neurodegeneration [81], pertinent given the cognitive impairments seen in many ME/CFS patients. Notably, our identification of a trans-pQTL for Factor H in *TNFRSF19*—a gene expressed in the central nervous system—suggests a link between peripheral complement regulation and neuroimmune signaling in ME/CFS. Ultimately, these findings position complement dysregulation not as an isolated feature but as a potential hallmark of chronic immunoinflammatory disease, offering a framework for therapeutic targeting and for refining biological subgroups across heterogeneous conditions.

In summary, our study gives an example of a pQTL-based approach integrating genotype with protein biomarker data that can uncover immune subtypes within a heterogeneous disease, offering a path forward for more personalized treatment strategies and mechanistic investigation. In addition, our study identifies inflammation driven by high C3 – likely resulting from increased inhibition of the alternative pathway – as a central feature in a genetically defined subgroup of ME/CFS. This subgroup, identifiable by complement pQTLs, may benefit from targeted therapies such as C3-lowering agents [82]. The complement abnormalities we observe mirror those in other IACCI, where imbalanced complement activity – whether genetically inherited or acquired – contributes to chronic inflammation and tissue damage. Our findings add to this landscape by highlighting the genetic variants in complement as a heritable contributor to ME/CFS susceptibility, where low basal alternative pathway activity may predispose individuals to persistent inflammation following an infectious or environmental insult. Complement also intersects with metabolic, endothelial, and coagulation pathways, linking dysregulation to hallmark ME/CFS symptoms like fatigue and bioenergetic stress [83].

The strengths of this study include use of participants from the population classified using a standardized approach and employing a novel pQTL analysis. Limitations include the relatively small number of ME/CFS patients, lack of racial diversity and cross-sectional data. Further study is required to evaluate the generalizability of our findings. Future studies should expand this approach genome- and proteome- wide and leverage resources like DecodeME or SearchME/CFS biorepository to explore stratification strategies across diverse populations. In parallel, tracking complement dynamics longitudinally – especially in patients with high-risk genotypes or post-infection – could clarify disease trajectories and inform therapeutic targeting.

## 4. Materials and Methods

### 4.1. Subject Recruitment and Characteristics

This study included the 50 ME/CFS and 121 non-fatigued (NF) controls subjects from the previously published genetic analysis focusing on inflammatory and immune-related pathways [3]. Briefly, participants were identified in the follow-up phase of the longitudinal study of ME/CFS in Georgia. ME/CFS cases were identified with the operationalized 1994 international research case definition [84]. Sample characteristics such as age, sex, body mass index (BMI), illness duration and onset, and symptom and functioning scores across fatigue (multidimensional fatigue inventory, MFI-20), functional status (SF-36), CDC symptom inventory (CDC-SI), and routine clinical tests results (including hs-CRP) were recorded for every participant [3]. The source study was approved by the Institutional Review Board of the Centers for Disease Control and Prevention and Abt Associates (now Abt Global). All subjects gave written informed consent for participating in the study and for anonymous testing and storage of biologic samples.

### 4.2. Plasma Protein and Complement Component Assays

Whole blood from each subject was collected in ethylenediaminetetraacetic acid (EDTA) tubes and processed for plasma within 30 minutes to minimize ex-vivo activation of complements. Plasma was separated, aliquoted, and stored at -70°C. A frozen plasma aliquot for each subject was shipped to the CLIA-certified Complement Laboratory at the National Jewish Health, Denver, Colorado for analysis of complement proteins/activation products using their validated clinical assays. Basic assay methods included measuring levels of C3a, C5a, Bb, SC5b-9 and Factor D by ELISA, C3 by nephelometry, and Factor B and Factor H by radial immunodiffusion. Results of hs-CRP testing were retrieved from the study database.

### 4.3. Statistical Analyses

#### 4.3.1. Data Transformation

Statistical and genetic analyses were performed on log<sub>2</sub>-transformed values of the measured proteins involved in the complement system (C3a, C5a, C3, Bb, SC5b-9, Factor D, Factor B, Factor H, and CRP). This data transformation was based on the observed variance and distribution of protein levels across all participants, performed using RStudio.

#### 4.3.2. Data analyses and visualizations for group phenotypes and complement protein levels

A Wilcoxon rank-sum test was used to test for significant differences in age and BMI between ME/CFS and NF control subjects. Radar plots made in Microsoft Excel were used to display the mean values for MFI-20 and SF-36 T-scores for both groups. A Pearson correlation test was performed to assess the association of complement factor levels with each other.

#### 4.3.3. Identification of Confounding Factors

A linear regression analysis was used to determine the bivariate association of circulating complement protein levels with demographic features (or variables) recognized to be associated with ME/CFS, such as age, sex, and BMI. Covariate adjustments specific to each circulating complement protein based on these associations and those supported by the literature were included in downstream analyses.

#### 4.3.4. Analysis of Circulating Complement Proteins Between ME/CFS and NF Subjects

Either binary or multinomial logistic regression analyses were performed to test whether individual complement proteins were significantly associated with group classification, maintaining covariate considerations. In the 2-group analysis, ME/CFS cases status ( $Y = 1$ ) vs. NF controls ( $Y=0$ ) was modeled as the binary outcome:

$$\text{logit}(P(\text{Subject} = \text{Total NF})) = \beta_0 + \beta_1(\log_2(\text{Protein levels})) + \beta_2(\text{Covariate}) \dots + \varepsilon$$

In the 4-group multinomial model, subgroup assignment ( $Y =$  one of the four groups) was used as the categorical outcome, with NFhet set as the reference category:

$$\log \frac{P(\text{Group} = k)}{P(\text{Group} = \text{NFhet})} = \beta_0 + \beta_1(\log_2(\text{Protein levels})) + \beta_2(\text{Covariate}) \dots + \varepsilon$$

For  $k \in \{\text{CFShet}, \text{CFSrem}, \text{NFrem}\}$

Likelihood ratio tests (LRTs) were performed to assess the overall contribution of each complement protein to the model. For each protein, we compared the full model (including the protein of interest as a predictor) against a reduced model (excluding the protein) using the following:

$$\text{LRT statistic} = -2 \times (\log \text{likelihood of reduced model} - \log \text{likelihood of full model})$$

Where the full model is:  $\text{Group} \sim \log_2(\text{Protein levels}) + \text{Covariates}$

And the null model is:  $\text{Group} \sim \text{Covariates}$

Pairwise linear regression analyses were performed to directly compare log<sub>2</sub>-transformed protein concentrations between groups while adjusting for covariates. These were applied to both the 2-group and 4-group comparisons:

$$\log_2(\text{Protein levels}) = \beta_0 + \beta_1(\text{Group}) + \beta_2(\text{Covariate}) \dots + \varepsilon$$

Group was treated as a binary or categorical variable, and pairwise p-values were extracted using estimated marginal means and post hoc contrasts to highlight specific group differences. Boxplots and dot graphs were created using the R package ggPlot2 to visualize protein distributions across groups.

#### 4.3.5. Analysis of Participant Survey Data with Circulating Complement Proteins

Linear regression modeling with previously noted covariate adjustments was performed to assess complement protein levels with participant survey data. Specifically, log<sub>2</sub>-transformed plasma protein levels were modeled as a function of continuous survey metrics (CDC-SI, MFI-20, and SF-36), using the formula:

$$\begin{aligned} \log_2(\text{Protein level}) & \\ &= \beta_0 + \beta_1(\text{functional/health score}) \\ &+ \beta_2(\text{Covariate}) \dots + \varepsilon \end{aligned}$$

RStudio was used to create a heatmap to display the regression  $\beta_1$  coefficient and the p-values for each association. To illustrate specific trends, representative correlation dot plots with corresponding R<sup>2</sup> values were created for analytes with significant associations, capturing the distribution of case- and control-level survey responses.

#### 4.3.6. Analysis of SNP Associations with Circulating Complement Proteins

A genotypic logistic regression using an additive model was previously applied to assess 9,146 SNPs that passed Affymetrix Human Immune and Inflammation Chip quality control and their associations with ME/CFS [3]. To investigate genetic variants for associations with complement components, we performed covariate-adjusted linear regression analyses in RStudio comparing full and reduced models for each SNP-protein pair. This analysis was conducted across all subjects, without stratification by illness group, to assess genetic regulation of complement protein expression. For each SNP-protein pair, a full linear model including the SNP genotype and relevant covariates was compared to a reduced model excluding the genotype.

Full model:

$$\begin{aligned} \log_2(\text{Protein level}) & \\ &= \beta_0 + \beta_1(\text{SNP genotype coded additively}) \\ &+ \beta_2(\text{Covariate}) \dots + \varepsilon \end{aligned}$$

Reduced model:

$$\log_2(\text{Protein level}) = \beta_0 + \beta_2(\text{Covariate}) \dots + \varepsilon$$

An F-statistic was calculated to compare the full and reduced models for each SNP-protein pair, testing whether the inclusion of genotype significantly improved model fit. The p-value from the full-versus-reduced (FvR) comparison was used to define significance. Additionally, Bonferroni correction was applied to account for multiple testing across all models, and these results are reported in Supplementary Tables 1-10. Allele frequencies, including the identification of major and minor alleles for the genetic analyses, were determined using data from this study population.

#### 4.3.7. Pathway Enrichment Analysis to Annotate SNP Associations with Circulating Complement Protein Levels

The linear regression analysis identified 3,192 SNPs significantly associated with circulating complement proteins (Full vs. Reduced Linear regression model,  $p < 0.05$ ). The 3,192 SNPs associated with circulating complement protein levels were further filtered to identify top associations (Full vs. Reduced Linear regression model,  $p < 0.01$ ) for each measured complement protein. This analysis resulted in 776 significant SNPs linked to complement protein plasma levels, representing a total of 359 unique genes. To investigate the distribution of significant SNPs across functional groups for each protein, data was extracted from Metascape, Golden Helix SVS software, and the literature to create a list of all genes with SNPs significantly associated with each complement protein and their function, allowing assignment of to one of seven functional groups (Table S11). For each complement protein, the number of significant genes and SNPs associated with each functional group was calculated. Enrichment scores were computed by dividing the observed proportion of genes/SNPs in a functional group within each complement factor's "top hits" list by the expected proportion of genes/SNPs in each functional group based on the overall distribution of significant genes/SNPs across all functional

groups and complement factor “top hits” lists. The overall distribution of significant genes/SNPs across all functional groups regardless of complement factor association was visualized by creating a donut plot using Microsoft Excel. Lollipop plots were generated using R to visualize these enrichment scores for each circulating complement protein measured, with the y-axis containing functional groups and overrepresentation in a group indicated by an enrichment score  $> 1$ .

#### 4.3.8. Analysis of SNP Associations with Circulating Complement Protein Levels and Disease Status

The 3,192 SNPs significantly associated with circulating complement proteins were compared against the previously published 32 functional SNPs, 10 proxy SNPs, and 6 additional pyro sequenced complement-linked SNPs significantly associated with ME/CFS (with sex and BMI covariate adjustments) [3], identifying 16 SNPs in common with both analyses and therefore associated with both disease status and circulating complement protein levels. These pQTLs were also assessed for whether they were cis or trans-acting using the QTLbase home database ([www.mulinlab.org](http://www.mulinlab.org)).

#### 4.3.9. Analysis of the Directionality of Genetic Variant Traits with Disease Risk

To investigate the broader pattern of disease-associated SNPs with complement protein levels, we assessed the relationship between the odds ratios (ORs) of significant genetic variants for ME/CFS risk and beta coefficients for circulating complement protein levels. ORs were derived from previously published SNP-disease associations (32 SNPs assessed against C3 beta values, and 27 SNPs assessed against Bb beta values) [3], and beta coefficients were generated from our circulating complement protein-level QTL analyses using the log<sub>2</sub>-transformed plasma concentrations of each complement protein. All SNPs previously reported as significantly associated with ME/CFS in the prior study [3] were included in this analysis, regardless of functional consequence. Specifically, we plotted ORs on the x-axis and corresponding beta values for C3 or Bb on the y-axis, with each point representing a single SNP significantly associated with ME/CFS status.

#### 4.3.10. Genotype-Stratified Analysis to Identify ME/CFS Subgroup Related to Complement System Dysregulation

The pQTLs identified in this study were evaluated for their ability to discriminate the illness status and subgrouping of the subjects in the study. This was done by first prioritizing pQTLs with high power for prediction of illness status by conducting a series of ROC analyses that split the subjects based on genotypes of QTLs, with log<sub>2</sub>-transformed plasma complement protein levels as the test variable. Specifically, participants were grouped by genotype at each SNP of interest, and logistic regression models were used to calculate predicted probabilities for ME/CFS status. These probabilities were then used to compute ROC curves and AUC values. Genotypes that resulted in better discrimination of subjects' illness status as indicated by high Area Under the Curve (AUC  $\geq 0.75$ ) were selected for combined analysis of markers to further improve the AUC for predicting illness status. An example of grouping subjects based on two markers rs9332739 and rs800292 was visualized to evaluate the discriminatory capacity of complement-associated pQTLs for potential subgroup identification within the study population. This combined marker analysis conducted using rs9332739 and rs800292 considered ME/CFS participants with heterozygous genotypes at either SNP as Subgroup A (CFS<sub>het</sub>), and the remainder (those without heterozygosity at either loci) were assigned to Subgroup B (CFS<sub>rem</sub>). Similarly, NF controls with heterozygous genotypes at either marker were assigned to Subgroup C (NF<sub>het</sub>), and all remaining controls were assigned to Subgroup D (NF<sub>rem</sub>). Of note, no individuals carried the homozygous risk genotype for rs9332739 (C/C), while five individuals carried the homozygous risk genotype for rs800292 (T/T). Because subgrouping was based on genotypes with the strongest predictive value and sufficiently robust sample sizes, two individuals (one ME/CFS case and one NF control) who were homozygous for the rs800292 risk allele were included in the broader CFS<sub>rem</sub> or NF<sub>rem</sub> groups, rather than in the heterozygous subgroup.

This four-group classification was applied to assess refined stratification of disease status and to characterize potential genotype-informed subgroups. Logistic models were rerun using these stratified groups to evaluate differences in complement protein levels, as described in 4.3.4., and ROC performance across identified subgroups.

#### 4.3.11. Comparison of Significant SNPs Between Our Population and the UK Biobank

To assess overlap between our top genetic variant associations and previously reported ME/CFS-related SNPs, we compared the top 50 SNP associations for each of the nine measured complement proteins with publicly available genome-wide association data from the UK biobank [85,86]. Specifically, we extracted SNPs significantly associated with ME/CFS-related phenotypes – including the groups Chronic Fatigue Syndrome, Post Viral Fatigue Syndrome, the broader R53 Fatigue and Malaise category, and Ever CFS (recovered ME/CFS subjects) – from the UK Biobank's large population-based prospective study. We then evaluated whether any of the top complement protein-associated SNPs in our dataset overlapped with those reported for these fatigue-related phenotypes in the UK biobank, and compared variant details, p-values, and odds ratios for common SNPs.

**Supplementary Materials:** The following supporting information can be found in the supplemental data files (Supplementary\_Materials\_Figures\_V4.docx, Suppl\_Tables\_V2.xlsx)

**Author Contributions:** Conceptualization: MSR, JM, and ERU; Methodology: MSR and JM; Software: MSR and JM; Validation: MSR and JM; Formal analysis: MSR and JM; Investigation: MSR and JM; Resources: MSR, JM, JSL, and ERU; Data curation: MSR, JSL, and JM; Writing—original draft preparation: MSR and JM; Writing—review and editing: MSR, JM, JSL, and ERU; Visualization: JM and MSR; Supervision: MSR and ERU; Project administration: MSR and ERU; Funding acquisition: MSR and ERU.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** As a requirement of publication, author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the authorship and conflict of interest criteria.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data supporting reported results can be found in supplemental materials. Researchers wanting to access the datasets used in this study should email CDC's ME/CFS Program (cfs@cdc.gov) and discuss next steps for the data request. The ME/CFS program data review committee will grant access after the review and the data use agreement is finalized.

**Acknowledgments:** Partial support for JM was provided by the research participation program at the Centers for Disease Control and Prevention (CDC), Division of High-Consequence Pathogens & Pathology, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the CDC. We thank Janna Murray for their invaluable assistance in acquiring and coordinating access to datasets used in this study. We are especially grateful to the patients and non-fatigue healthy controls who generously participated in the research, making this work possible. Their contributions continue to advance our understanding of ME/CFS and related conditions.

**Conflicts of Interest:** The findings and the conclusions in this report are those of the authors and do not necessarily represent the official position of CDC. Authors declare no conflict of interest.

## References

1. Medicine, I.o., *Beyond Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Redefining an Illness*. 2015, Washington, DC: The National Academies Press. 304.

2. Unger, E.R., et al., *Heterogeneity in Measures of Illness among Patients with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Is Not Explained by Clinical Practice: A Study in Seven U.S. Specialty Clinics*. Journal of Clinical Medicine, 2024. **13**(5): p. 1369.
3. Rajeevan, M.S., et al., *Pathway-focused genetic evaluation of immune and inflammation related genes with chronic fatigue syndrome*. Hum Immunol, 2015. **76**(8): p. 553-60.
4. Smith, A.K., et al., *Convergent genomic studies identify association of GRIK2 and NPAS2 with chronic fatigue syndrome*. Neuropsychobiology, 2011. **64**(4): p. 183-94.
5. Perez, M., et al., *Genetic Predisposition for Immune System, Hormone, and Metabolic Dysfunction in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: A Pilot Study*. Front Pediatr, 2019. **7**: p. 206.
6. Schlauch, K.A., et al., *Genome-wide association analysis identifies genetic variations in subjects with myalgic encephalomyelitis/chronic fatigue syndrome*. Transl Psychiatry, 2016. **6**(2): p. e730.
7. Marshall-Gradisnik, S., et al., *Natural killer cells and single nucleotide polymorphisms of specific ion channels and receptor genes in myalgic encephalomyelitis/chronic fatigue syndrome*. Appl Clin Genet, 2016. **9**: p. 39-47.
8. Marshall-Gradisnik, S., et al., *Single nucleotide polymorphisms and genotypes of transient receptor potential ion channel and acetylcholine receptor genes from isolated B lymphocytes in myalgic encephalomyelitis/chronic fatigue syndrome patients*. J Int Med Res, 2016. **44**(6): p. 1381-1394.
9. Carlo-Stella, N., et al., *A first study of cytokine genomic polymorphisms in CFS: Positive association of TNF-857 and IFN $\gamma$  874 rare alleles*. Clin Exp Rheumatol, 2006. **24**(2): p. 179-82.
10. Steiner, S., et al., *Autoimmunity-Related Risk Variants in PTPN22 and CTLA4 Are Associated With ME/CFS With Infectious Onset*. Front Immunol, 2020. **11**: p. 578.
11. Lande, A., et al., *Human Leukocyte Antigen alleles associated with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS)*. Sci Rep, 2020. **10**(1): p. 5267.
12. Rajeevan, M.S., et al., *Glucocorticoid receptor polymorphisms and haplotypes associated with chronic fatigue syndrome*. Genes Brain Behav, 2007. **6**(2): p. 167-76.
13. Tziastoudi, M., et al., *Genetics of COVID-19 and myalgic encephalomyelitis/chronic fatigue syndrome: a systematic review*. Ann Clin Transl Neurol, 2022. **9**(11): p. 1838-1857.
14. Hajdarevic, R., et al., *Genetic association study in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) identifies several potential risk loci*. Brain Behav Immun, 2022. **102**: p. 362-369.
15. Yao, C., et al., *Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease*. Nat Commun, 2018. **9**(1): p. 3268.
16. Zhao, J.H., et al., *Genetics of circulating inflammatory proteins identifies drivers of immune-mediated disease risk and therapeutic targets*. Nat Immunol, 2023. **24**(9): p. 1540-1551.
17. He, B., et al., *Genome-wide pQTL analysis of protein expression regulatory networks in the human liver*. BMC Biol, 2020. **18**(1): p. 97.
18. Ricklin, D., et al., *Complement: a key system for immune surveillance and homeostasis*. Nature Immunology, 2010. **11**(9): p. 785-797.
19. Mathern, D.R. and P.S. Heeger, *Molecules Great and Small: The Complement System*. Clin J Am Soc Nephrol, 2015. **10**(9): p. 1636-50.
20. Sorensen, B., et al., *Complement activation in a model of chronic fatigue syndrome*. J Allergy Clin Immunol, 2003. **112**(2): p. 397-403.
21. Sorensen, B., et al., *Transcriptional control of complement activation in an exercise model of chronic fatigue syndrome*. Mol Med, 2009. **15**(1-2): p. 34-42.
22. Giloteaux, L., et al., *Dysregulation of extracellular vesicle protein cargo in female myalgic encephalomyelitis/chronic fatigue syndrome cases and sedentary controls in response to maximal exercise*. J Extracell Vesicles, 2024. **13**(1): p. e12403.
23. Schutzer, S.E., et al., *Distinct cerebrospinal fluid proteomes differentiate post-treatment lyme disease from chronic fatigue syndrome*. PLoS One, 2011. **6**(2): p. e17287.
24. Germain, A., S.M. Levine, and M.R. Hanson, *In-Depth Analysis of the Plasma Proteome in ME/CFS Exposes Disrupted Ephrin-Eph and Immune System Signaling*. Proteomes, 2021. **9**(1).
25. Guenther, S., et al., *Frequent IgG subclass and mannose binding lectin deficiency in patients with chronic fatigue syndrome*. Human Immunology, 2015. **76**(10): p. 729-735.

26. Lutz, L., et al., *Evaluation of Immune Dysregulation in an Austrian Patient Cohort Suffering from Myalgic Encephalomyelitis/Chronic Fatigue Syndrome*. *Biomolecules*, 2021. **11**(9).
27. Ding, X., M. Patel, and C.C. Chan, *Molecular pathology of age-related macular degeneration*. *Prog Retin Eye Res*, 2009. **28**(1): p. 1-18.
28. Despret, D.D., et al., *Complement factor H polymorphism, complement activators, and risk of age-related macular degeneration*. *JAMA*, 2006. **296**(3): p. 301-9.
29. Gold, B., et al., *Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration*. *Nat Genet*, 2006. **38**(4): p. 458-62.
30. Scholl, H.P., et al., *Systemic complement activation in age-related macular degeneration*. *PLoS One*, 2008. **3**(7): p. e2593.
31. Hecker, L.A., et al., *Genetic control of the alternative pathway of complement in humans and age-related macular degeneration*. *Hum Mol Genet*, 2010. **19**(1): p. 209-15.
32. Smailhodzic, D., et al., *Risk alleles in CFH and ARMS2 are independently associated with systemic complement activation in age-related macular degeneration*. *Ophthalmology*, 2012. **119**(2): p. 339-46.
33. Sebastián-Martín, A., et al., *Role of Dipeptidyl Peptidase-4 (DPP4) on COVID-19 Physiopathology*. *Biomedicines*, 2022. **10**(8).
34. Hoke, M., et al., *Polymorphism of the complement 5 gene and cardiovascular outcome in patients with atherosclerosis*. *Eur J Clin Invest*, 2012. **42**(9): p. 921-6.
35. Giles, J.L., et al., *Functional analysis of a complement polymorphism (rs17611) associated with rheumatoid arthritis*. *J Immunol*, 2015. **194**(7): p. 3029-34.
36. Acar, I.E., et al., *Semi-Quantitative Multiplex Profiling of the Complement System Identifies Associations of Complement Proteins with Genetic Variants and Metabolites in Age-Related Macular Degeneration*. *J Pers Med*, 2021. **11**(12).
37. Copenhaver, M.M., et al., *Relationships of complement components C3 and C4 and their genetics to cardiometabolic risk in healthy, non-Hispanic white adolescents*. *Pediatr Res*, 2020. **87**(1): p. 88-94.
38. Gaya da Costa, M., et al., *Age and Sex-Associated Changes of Complement Activity and Complement Levels in a Healthy Caucasian Population*. *Front Immunol*, 2018. **9**: p. 2664.
39. LaFon, D.C., et al., *Classical and lectin complement pathways and markers of inflammation for investigation of susceptibility to infections among healthy older adults*. *Immunity & Ageing*, 2020. **17**(1): p. 18.
40. Kelkar, N.S., et al., *Sex- and species-associated differences in complement-mediated immunity in humans and rhesus macaques*. *mBio*, 2024. **15**.
41. Butler, A.E., et al., *Complement Dysregulation in Obese Versus Nonobese Polycystic Ovary Syndrome Patients*. *Cells*, 2023. **12**(15).
42. Qiao, Q., et al., *Glucose Restriction Plus Refeeding in Vitro Induce Changes of the Human Adipocyte Secretome with an Impact on Complement Factors and Cathepsins*. *Int J Mol Sci*, 2019. **20**(16).
43. Fruh, S.M., *Obesity: Risk factors, complications, and strategies for sustainable long-term weight management*. *J Am Assoc Nurse Pract*, 2017. **29**(S1): p. S3-s14.
44. Moreno-Navarrete, J.M. and J.M. Fernández-Real, *The complement system is dysfunctional in metabolic disease: Evidences in plasma and adipose tissue from obese and insulin resistant subjects*. *Semin Cell Dev Biol*, 2019. **85**: p. 164-172.
45. Mawle, A.C., et al., *Immune responses associated with chronic fatigue syndrome: a case-control study*. *J Infect Dis*, 1997. **175**(1): p. 136-41.
46. Valdez, A.R., et al., *Estimating Prevalence, Demographics, and Costs of ME/CFS Using Large Scale Medical Claims Data and Machine Learning*. *Front Pediatr*, 2018. **6**: p. 412.
47. Olivieri, F., et al., *Sex/gender-related differences in inflammaging*. *Mech Ageing Dev*, 2023. **211**: p. 111792.
48. Kamitaki, N., et al., *Complement genes contribute sex-biased vulnerability in diverse disorders*. *Nature*, 2020. **582**(7813): p. 577-581.
49. Sodhi, E.U., et al., *Sex-Differences and Associations Between Complement Activation and Synovial Vascularization in Patients with Late-Stage Knee Osteoarthritis*. *Front Immunol*, 2022. **13**: p. 890094.
50. Rohrhofer, J., et al., *Immunological Patient Stratification in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome*. *J Clin Med*, 2024. **13**(1).

51. Castro-Marrero, J., et al., *Complement Component C1q as a Potential Diagnostic Tool for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Subtyping*. J Clin Med, 2021. **10**(18).
52. Buchwald, D., et al., *Markers of inflammation and immune activation in chronic fatigue and chronic fatigue syndrome*. J Rheumatol, 1997. **24**(2): p. 372-6.
53. Nacul, L., et al., *Evidence of Clinical Pathology Abnormalities in People with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) from an Analytic Cross-Sectional Study*. Diagnostics (Basel), 2019. **9**(2).
54. Nunes, M., et al., *Data-independent LC-MS/MS analysis of ME/CFS plasma reveals a dysregulated coagulation system, endothelial dysfunction, downregulation of complement machinery*. Cardiovasc Diabetol, 2024. **23**(1): p. 254.
55. Paun, C.C., et al., *A Novel Complotype Combination Associates with Age-Related Macular Degeneration and High Complement Activation Levels in vivo*. Sci Rep, 2016. **6**: p. 26568.
56. Loirat, C. and V. Frémeaux-Bacchi, *Atypical hemolytic uremic syndrome*. Orphanet J Rare Dis, 2011. **6**: p. 60.
57. Bogdan, R.G., et al., *Atypical Hemolytic Uremic Syndrome: A Review of Complement Dysregulation, Genetic Susceptibility and Multiorgan Involvement*. J Clin Med, 2025. **14**(7).
58. Tortajada, A., et al., *The disease-protective complement factor H allotypic variant Ile62 shows increased binding affinity for C3b and enhanced cofactor activity*. Hum Mol Genet, 2009. **18**(18): p. 3452-61.
59. Lucientes-Contente, L., B. Márquez-Tirado, and E. Goicoechea de Jorge, *The Factor H protein family: The switchers of the complement alternative pathway*. Immunological Reviews, 2022. **313**: p. 25 - 45.
60. Chen, S.F., et al., *Plasma complement factor H is associated with disease activity of patients with ANCA-associated vasculitis*. Arthritis Res Ther, 2015. **17**(1): p. 129.
61. Salminen, A., et al., *Genetic Variants Contributing to Circulating Matrix Metalloproteinase 8 Levels and Their Association With Cardiovascular Diseases: A Genome-Wide Analysis*. Circ Cardiovasc Genet, 2017. **10**(6).
62. Cheng, J., et al., *Regulatory Architecture of the RCA Gene Cluster Captures an Intragenic TAD Boundary, CTCF-Mediated Chromatin Looping and a Long-Range Intergenic Enhancer*. Front Immunol, 2022. **13**: p. 901747.
63. Tate, W., et al., *Molecular Mechanisms of Neuroinflammation in ME/CFS and Long COVID to Sustain Disease and Promote Relapses*. Front Neurol, 2022. **13**: p. 877772.
64. Komaroff, A.L. and W.I. Lipkin, *ME/CFS and Long COVID share similar symptoms and biological abnormalities: road map to the literature*. Front Med (Lausanne), 2023. **10**: p. 1187163.
65. Gil, A., et al., *Identification of CD8 T-cell dysfunction associated with symptoms in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) and Long COVID and treatment with a nebulized antioxidant/anti-pathogen agent in a retrospective case series*. Brain, Behavior, & Immunity - Health, 2024. **36**: p. 100720.
66. Iu, D.S., et al., *Transcriptional reprogramming primes CD8+ T cells toward exhaustion in Myalgic encephalomyelitis/chronic fatigue syndrome*. Proceedings of the National Academy of Sciences, 2024. **121**(50): p. e2415119121.
67. Lage, S.L., et al., *Persistent immune dysregulation and metabolic alterations following SARS-CoV-2 infection*. medRxiv, 2025.
68. Mandarano, A.H., et al., *Myalgic encephalomyelitis/chronic fatigue syndrome patients exhibit altered T cell metabolism and cytokine associations*. The Journal of Clinical Investigation, 2020. **130**(3): p. 1491-1505.
69. Cervia-Hasler, C., et al., *Persistent complement dysregulation with signs of thromboinflammation in active Long Covid*. Science, 2024. **383**(6680): p. eadg7942.
70. Baillie, K., et al., *Complement dysregulation is a prevalent and therapeutically amenable feature of long COVID*. Med, 2024. **5**(3): p. 239-253.e5.
71. Ahearn-Ford, S., et al., *Long-term disruption of cytokine signalling networks is evident in patients who required hospitalization for SARS-CoV-2 infection*. Allergy, 2021. **76**(9): p. 2910-2913.
72. Lai, Y.-J., et al., *Biomarkers in long COVID-19: A systematic review*. Frontiers in Medicine, 2023. **Volume 10 - 2023**.
73. Hagiya, H., et al., *Relevance of complement immunity with brain fog in patients with long COVID*. J Infect Chemother, 2024. **30**(3): p. 236-241.
74. Shih, P.B., et al., *Genetic variation in C-reactive protein (CRP) gene may be associated with risk of systemic lupus erythematosus and CRP concentrations*. J Rheumatol, 2008. **35**(11): p. 2171-8.

75. Zhao, J., et al., *Association of genetic variants in complement factor H and factor H-related genes with systemic lupus erythematosus susceptibility*. PLoS Genet, 2011. 7(5): p. e1002079.
76. Kelley, J.M., J.C. Edberg, and R.P. Kimberly, *Pathways: Strategies for susceptibility genes in SLE*. Autoimmun Rev, 2010. 9(7): p. 473-6.
77. Berndtson, K., *Review of evidence for immune evasion and persistent infection in Lyme disease*. Int J Gen Med, 2013. 6: p. 291-306.
78. Hovis, K.M., et al., *Selective binding of Borrelia burgdorferi OspE paralogs to factor H and serum proteins from diverse animals: possible expansion of the role of OspE in Lyme disease pathogenesis*. Infect Immun, 2006. 74(3): p. 1967-72.
79. Parnell, G.P. and D.R. Booth, *The Multiple Sclerosis (MS) Genetic Risk Factors Indicate both Acquired and Innate Immune Cell Subsets Contribute to MS Pathogenesis and Identify Novel Therapeutic Opportunities*. Front Immunol, 2017. 8: p. 425.
80. Ingram, G., et al., *Complement in multiple sclerosis: its role in disease and potential as a biomarker*. Clin Exp Immunol, 2009. 155(2): p. 128-39.
81. Negro-Demontel, L., et al., *The complement system in neurodegenerative and inflammatory diseases of the central nervous system*. Front Neurol, 2024. 15: p. 1396520.
82. Lamers, C., et al., *Compstatins: the dawn of clinical C3-targeted complement inhibition*. Trends Pharmacol Sci, 2022. 43(8): p. 629-640.
83. Nunes, J.M., D.B. Kell, and E. Pretorius, *Cardiovascular and haematological pathology in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS): A role for viruses*. Blood Rev, 2023. 60: p. 101075.
84. Reeves, W.C., et al., *Prevalence of chronic fatigue syndrome in metropolitan, urban, and rural Georgia*. Population Health Metrics, 2007. 5(1): p. 5.
85. Sudlow, C., et al., *UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age*. PLoS Med, 2015. 12(3): p. e1001779.
86. Wang, Q., et al., *Rare variant contribution to human disease in 281,104 UK Biobank exomes*. Nature, 2021. 597(7877): p. 527-532.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.