BMC Infectious Diseases

RESEARCH

Open Access

# A comparison of genome-wide association analyses of persistent symptoms after Lyme disease, fibromyalgia, and myalgic encephalomyelitis – chronic fatigue syndrome

Annemarie G. Hirsch[1,2*], Anne E. Justice[1], Amy Poissant[1], Cara M. Nordberg[1], Navya S. Josyula[1], John Aucott[2], Alison W. Rebman[2] and Brian S. Schwartz[1,3]

## Abstract

**Background** Up to 20% of Lyme disease cases experience post-treatment Lyme disease syndrome (PTLDS). The biological basis for PTLDS is poorly understood and no evidence-based treatment has been identified. Genetic studies have the potential to elucidate PTLDS pathophysiology and identify treatment targets.

**Methods** We used electronic health record data (EHR) and genetic data from a linked biorepository to conduct a genome-wide association study (GWAS) for PTLDS among patients from a Pennsylvania health system. We evaluated the validity of the GWAS results in two separate conditions that have hypothesized overlapping pathophysiology, fibromyalgia and myalgic encephalomyelitis – chronic fatigue syndrome (ME/CFS). GWAS analyses were performed using logistic regression in SUGEN, assuming an additive genetic model, and adjusting for age, sex, array, and the first 10 principal components calculated from whole genome genotyping to adjust for ancestry, and accounting for relatedness including all 1st degree relationships. The functional mapping and annotation analysis (FUMA) tool was used to explore top findings from our GWAS.

**Results** Among the 161,875 eligible MyCode participants with genotyping, there were 3,585 who met the criteria for treated Lyme disease. A subset of 695 (19.4%) of these patients met the criteria for PTLDS and the remaining 2890 were classified as controls. We identified two PTLDS loci that reached the suggestive significance threshold ($P < 5 \times 10^{-7}$), with lead variants rs77857587, near *IRX1*, and rs10833979, near *GAS2*. Our top index single nucleotide polymorphism (SNP), rs77857587, is in high linkage disequilibrium with a long-range protein quantitative locus SNP, rs111774530, for the MARC2 (Mitochondrial Amidoxime Reducing Component 2) protein. We identified 5,041 cases of fibromyalgia (150,599 controls) and 2,268 cases of ME/CFS (151,594 controls) among the MyCode participants. Neither of the two suggestively significant loci were associated with fibromyalgia or ME/CFS.

**Conclusion** We identified two PTLDS loci that reached a suggestive significance threshold. Our top index SNP is associated with the MARC2 protein, a protein that has been linked to multiple immune checkpoints. Further study is

*Correspondence:
Annemarie G. Hirsch
aghirsch@geisinger.edu

Full list of author information is available at the end of the article

needed in a larger population to evaluate whether there is genetic evidence of the role of immune response in the occurrence of PTLDS.

**Keywords** Lyme disease, Fibromyalgia, Myalgic encephalomyelitis, Genome-wide association study

## Introduction

An estimated 465,000 Americans are treated for Lyme disease annually [1]. Within the United States, Lyme disease primarily occurs in the Northeast and mid-Atlantic regions and in the upper Midwest, with additional focal areas of transmission along the West Coast [1]. However, the geography of Lyme disease risk has expanded over the last two decades in all directions into neighboring states and adjacent Canada, where Lyme disease is a rapidly emerging public health problem [1]. While Lyme disease usually responds well to antibiotic therapy [2], up to 20% of cases experience persistent symptoms after antibiotic treatment and meet criteria for what has been defined as post-treatment Lyme disease syndrome (PTLDS) [3]. PTLDS is characterized by persistent or recurrent symptoms, lasting six months or more, of fatigue, musculoskeletal pain, and cognitive complaints leading to decline in physical and social functioning [4, 5]. The biological basis for PTLDS is poorly understood and no evidence-based treatment has been identified [5].

There are a few competing hypotheses regarding the pathogenesis of PTLDS, making it a challenge to prevent, treat, and target further research. PTLDS may be due to an ongoing host inflammatory response, independent of ongoing infection [6]. Alternatively, PTLDS may be driven by an inability to eliminate the infection [2, 6]. However, randomized trials of antibiotics to treat PTLDS have not demonstrated clear benefit [4, 7–9]. Finally, PTLDS could be the result of central sensitization, a process of hyperactivation in the central neural pathways, leading to a more intense response to sensory stimuli [10]. Conditions with overlapping symptoms of PTLDS, including fibromyalgia and myalgic encephalomyelitis (also known as chronic fatigue syndrome, ME/CFS), have been linked to central sensitization syndrome (CSS) [11].

Genetic studies have the potential to elucidate the pathophysiology of PTLDS. To date, there have been no studies of the genetic architecture contributing to the risk of PTLDS. Limited research has examined the genetics of Lyme disease itself. In mouse and human studies, genetic polymorphisms have been associated with different inflammatory responses to *B. burgdorferi* [12–14]. Genetic studies have not explored the multiple hypothesized pathways to PTLDS.

Elucidating the pathogenesis of PTLDS is a critical step to identifying treatments that can prevent PTLDS or bring relief to individuals suffering from PTLDS. We used electronic health record data (EHR) and genetic data from the MyCode™ Community Health Initiative biorepository (MyCode) to conduct a genome-wide association study (GWAS) for PTLDS among patients from a health system serving central and northeastern Pennsylvania, a region highly endemic to Lyme disease. We evaluated the validity of the GWAS results in two separate cohorts with conditions that have hypothesized overlapping pathophysiology, fibromyalgia and ME/CFS.

## Methods

### Study population

This study was conducted among adult patients (18 years of age and older) of Geisinger, a large integrated health system in Pennsylvania. Participants had to have consented to MyCode, a system-wide biorepository of blood, serum, and DNA samples at Geisinger. MyCode has been enrolling Geisinger patients since early 2007, with a consent rate of more than 85% [15] Eligibility criteria is not dependent on a particular health condition and recruitment occurs across diverse primary and specialty care clinics, providing a reasonable approximation of a random sample of the Geisinger adult patient population [15]. MyCode samples and genetic data can be linked to information in participants' EHRs. This study included all patients with treated Lyme disease documented in the EHR who had their DNA sequenced for MyCode.

### Identification of PTLDS cases

We identified PTLDS cases and controls from among individuals diagnosed and treated for Lyme disease. To identify individuals with Lyme disease in the EHR we applied a previously described approach [16]. Individuals had to have either a diagnosis for Lyme disease or a serology Current Procedural Terminology (CPT) order (enzyme immunoassay or Western blot for immunoglobulin M or G) code. For diagnoses we used both Epic (Verona, WI) electronic diagnosis group (EDG) names or International Classification of Diseases Ninth and Tenth Revision Clinical Modification (ICD-9-CM 088.81 and ICD-10-CM A69.2.2x) codes. Individuals must also have had an order for an antibiotic within 30 days of the date of the diagnosis or the serology test order. To improve the specificity of this definition, antibiotic orders were excluded if linked to respiratory disease or otitis media, since these are common diagnoses treated with the same antibiotics as Lyme disease.

Among individuals classified as having treated Lyme disease, we confined the study sample to individuals with at least 26 weeks of observation in the EHR before Lyme disease and 52 weeks of observation after, to have

sufficient observation time to detect the absence or presence of symptoms consistent with PTLDS. We then identified individuals with evidence of PTLDS using a previously described definition: [16] At least one diagnosis linked to an encounter or medication order within three categories of symptoms: malaise or fatigue, pain, or cognitive difficulties that was dated in the EHR four to 52 weeks after Lyme disease, with no history of the diagnosis code four to 26 weeks prior to Lyme disease. To increase the sensitivity of this definition, we added two additional inclusion criteria. First, we also classified individuals as having PTLDS if they had a diagnosis for PTLDS associated with a clinical encounter. Second, we created an algorithm to capture free-text notes indicating PTLDS. For the latter, we conducted interviews with clinicians in infectious disease, internal medicine, and family medicine to identify commonly used terms for documenting PTLDS in EHR clinician notes. Based on these interviews, we selected five terms: PTLDS, post-viral syndrome, post-Lyme syndrome, Lyme syndrome, or post-viral. We then conducted an automated search for these terms, removing negated terms (e.g., "no post-viral syndrome"), in all of the free-text clinician notes between four to 52 weeks after meeting the criteria for Lyme disease. In sensitivity analyses, we applied a more restrictive definition to PTLDS that required evidence of symptoms during a shorter look-back period of four to 26 weeks.

#### Identification of fibromyalgia and ME/CFS cases

To examine potential shared genetic associations with similar phenotypes, we performed secondary analyses on suggestively significant ($P < 5 \times 10^{-7}$) PTLDS loci with fibromyalgia or ME/CFS. Individuals were classified as having the condition if they had at least two diagnosis codes for clinical encounters on two different dates at any time in the EHR. Fibromyalgia was identified using the ICD-10-CM code M79.1 or a linked EDG diagnoses. ME/CFS was identified using ICD-9-CM 780.71, ICD-10-CM R53.82, or a linked EDG diagnoses code.

#### Controls

For PTLDS, controls were defined as MyCode participants with sequenced DNA with treated Lyme disease who did not meet the criteria for PTLDS and who had data in the EHR for at least 26 weeks prior to meeting treated Lyme disease criteria. For fibromyalgia and ME/CFS, controls were defined as MyCode participants with sequenced DNA who did not meet the case definitions and did not have any fibromyalgia or ME/CFS diagnoses, respectively, in the EHR at any time.

#### Genotyping and imputation of genetic data

Each participant was genotyped on the Illumina GSA (Global Sequencing Array) v1 or v2, or the Illumina OmniExpressExome v1.2, v1.3, or v1.4 arrays. Quality control of the genetic data was performed by array and self-identified race and ethnicity, including sex concordance, single nucleotide polymorphism (SNP) and sample call rate filters (95%), identity by descent (IBD) (confirmed relatedness and performed across race/ethnicity), and Hardy-Weinberg equilibrium (HWE) ($p < 1 \times 10^{-15}$). Following initial quality control, race and ethnic groups were merged within array and call rate filters were applied again (90%). Quality controlled genotyping data were imputed by array to the NHLBI Trans-Omics for Precision Medicine (TOPMed) program, Release 2, reference panel using the TOPMed Imputation server [17–19]. The resulting imputation was merged across batch and filtered based on imputation quality (mean Rsq > 0.7), minor allele count greater than five (MAC > 5), and variant-level missingness < 5% (call rate > 95%).

#### Association analyses

All GWAS analyses were performed using logistic regression in SUGEN [20], assuming an additive genetic model, and adjusting for age, sex, array, and the first 10 principal components calculated from whole genome genotyping to adjust for ancestry, and accounting for relatedness within family networks including all 1st degree relationships using generalized estimating equations (GEEs) as implemented in SUGEN. The program PRIMUS (Pedigree Reconstruction and Identification of Maximally Unrelated Set) was used to infer these family networks/pedigrees using genome-wide data using genome-wide common and independent variants [21], as previously described in more detail [22]. All GWAS results were examined for signs of systematic errors (i.e., strand flips, missing or invalid summary statistics, inflation) following standard protocols and filtered for common variants (minor allele frequency [MAF] > 1%) and per-SNP minimum sample size of 30 using the R package EasyQC [23]. Due to a small proportion of non-European American participants, we performed genetic association analyses both across all race and ethnic groups and restricted to only self-identified White/European Americans to assess potential heterogeneity in our PTLDS GWAS. We used well-established thresholds for genome-wide significance with SNPs considered significantly associated with a $p < 5 \times 10^{-8}$ and suggestively significant at a threshold of $p < 5 \times 10^{-7}$. For robustness in the sensitivity analyses of more restrictive phenotype, we require P to remain within one order of magnitude of suggestive significance ($P < 5 \times 10^{-6}$) and effect estimate (Beta) remains > 90% of original effect and directionally consistent.

The functional mapping and annotation (FUMA) analysis tool was used to further explore top findings from our GWAS summary statistics. The default settings were selected for these analyses. Specifically, the SNP2GENE function was used to perform functional annotation for SNPs in significant loci, identify previously published GWAS associations nearby, and produce regional association plots [24, 25].

## Results

### PTLDS cases

Among the 161,875 eligible MyCode participants with existing genotypes, there were 3,585 patients who met the criteria for treated Lyme disease. A subset of 695 (19.4%) of the patients treated for Lyme disease met the criteria for PTLDS and the remainder were classified as controls (Fig. 1). Of the 695 PTLDS cases identified, 638 had a diagnosis consistent with PTLDS symptoms following the Lyme disease treatment: 57.7% had a diagnosis related to fatigue, 34% had a diagnosis related to pain, and 13.2% had a diagnosis related to cognitive symptoms (Table 1). Only a small proportion of these cases had an EDG diagnosis for PTLDS (1.3%) or a mention of a PTLDS-related term in clinician free-text notes (7.2%). Of the 100 cases who had PTLDS noted in clinician free-text notes, more than half did not have a diagnosis related to PTLDS symptoms. Only fourteen cases had an EDG diagnosis for PTLDS.

Following the previously described quality control, our genetic analysis of PTLDS included 670 PTLDS cases and 2653 controls. In comparing cases to controls, there was a small, but significant difference in the mean age, 1.5 years younger in cases ($p = 0.019$, Table 2). Additionally, we observed a higher proportion of women as compared to men among PTLDS cases ($p = 4 \times 10^{-4}$). We did not observe significant differences in the distribution of race/ethnic groups across cases and controls except for Hispanic/Latino participants, who were more common among PTLDS cases than controls ($p = 3 \times 10^{-4}$).

In our analysis that included all participants, we identified two PTLDS loci that reached the suggestive significance threshold ($P < 5 \times 10^{-7}$), one with lead SNP, rs77857587, that lies approximately 252 kb upstream of the *IRX1* (Iroquois Homeobox 1) gene and lead SNP, rs10833979, 325 kb downstream of *GAS2* (Growth Arrest Specific 2) (Table 3; Fig. 2, Additional file 1). Both suggestive loci ($P < 5 \times 10^{-7}$) were also identified in our GWAS restricted to White/European American participants, and no additional loci were identified (Table 3, Additional file 1). The lead variant for the locus near *GAS2* remained the same in this sub-analysis. However, the lead variant for the locus near *IRX1* changed in the White/European American-only analysis from rs77857587 to rs62337498, although to a nearby variant in high LD (distance = 13,312 bp, $R^2 = 0.958$). Thus, given the overall association signals remain the same, only results from our pooled GWAS analysis are discussed further. We did not identify any PTLDS loci that reached the genome-wide significance threshold of $p < 5 \times 10^{-8}$. None of our suggestively significant ($P < 5 \times 10^{-7}$) SNPs had a clear functional role related to nearby genes as indicated by our FUMA analysis. For example, all SNPs were intergenic or intronic, had CADD score [26] <20 and non-significant evidence of regulatory role as noted by RegulomeDB [27] scores ranging from 4 (minimal binding evidence) to 7 (no evidence). Of note, our top index SNP, rs77857587, is in high linkage disequilibrium (LD, $R^2 = 1.0$ in 1000 Genomes EUR superfamily and $R^2 = 0.7215$ in all 1000 Genomes samples) with a long-range protein quantitative locus (pQTL) SNP, rs111774530, for the MARC2 (Mitochondrial Amidoxime Reducing Component 2) protein [28–30]. When the more restrictive definition for PTLDS was applied, 121 cases were reclassified, and excluded from both cases and controls. However, the results of the association analyses did not change significantly and remain robust (**Table 3**, $P < 5 \times 10^{-6}$ and >90% original Beta).

### Secondary analyses

We identified 5,041 cases of fibromyalgia (150,599 controls) and 2,268 cases of ME/CFS (151,594 controls) among the MyCode participants (Table 2). Similar to PTLDS, cases tended to be younger than controls for fibromyalgia, on average, 1.37 years ($p = 1.63 \times 10^{-8}$); however, no significant difference in age was observed for ME/CFS. Additionally, there was a higher proportion of women compared to men among cases for both fibromyalgia and ME/CFS ($p < 4.5 \times 10^{-308}$ and $1.06 \times 10^{-14}$, respectively). For fibromyalgia, there was a lower proportion of White/European American participants ($p = 4.1 \times 10^{-4}$) and a higher proportion of Hispanic/Latino participants ($p = 6.5 \times 10^{-8}$) among cases as compared to controls. In contrast there were significantly fewer African American/Black participants and Hispanic/Latino participants ($p = 9.9 \times 10^{-3}$ and 0.0122, respectively) and a higher proportion of white/European American participants ($p = 2.2 \times 10^{-3}$) among ME/CFS cases. Neither of the two suggestively significant ($P < 5 \times 10^{-7}$) loci were associated with fibromyalgia or ME/CFS (both $p > 0.05$, **Table 3**).

## Discussion

We conducted a GWAS of PTLDS among adults residing in a region highly endemic for Lyme disease. To our knowledge, this is the first GWAS of PTLDS and the first comparison GWAS of three common conditions with overlapping symptom profiles. We leveraged discrete and free-text data fields in the EHR to identify PTLDS cases

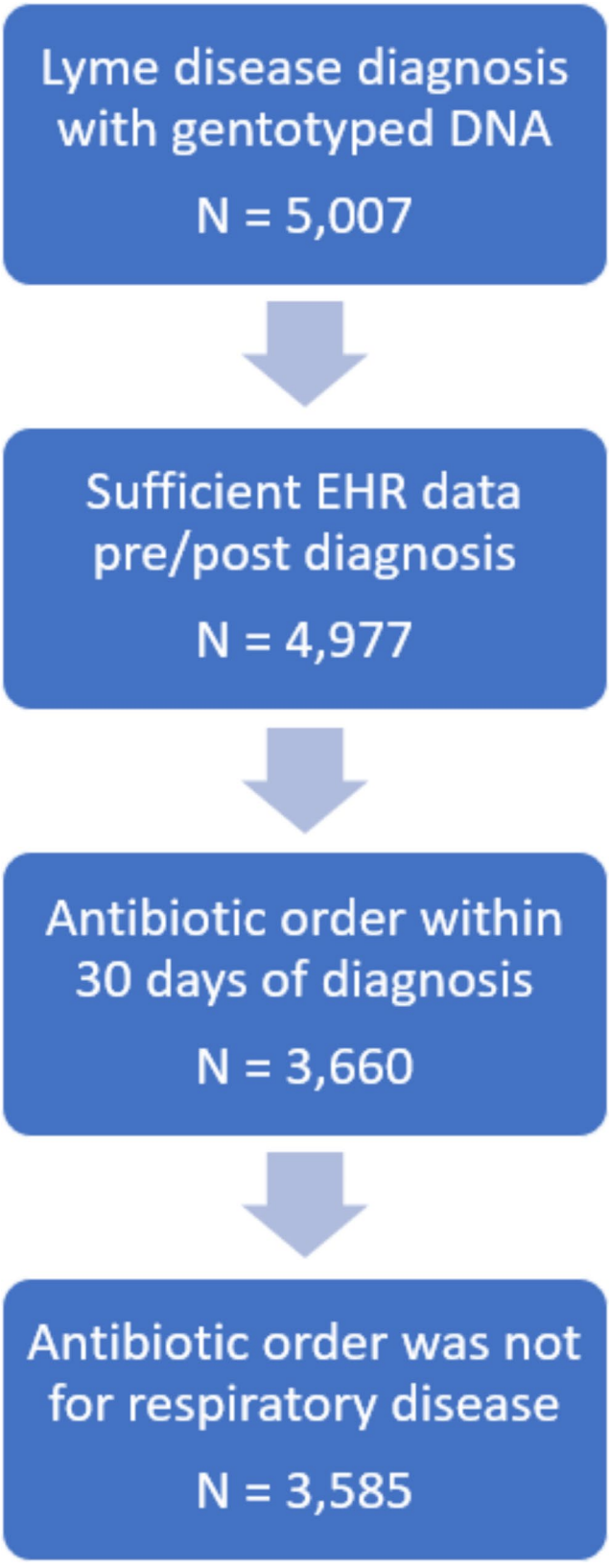


**Fig. 1** Process flow for identifying cases of treated Lyme disease. Abbreviations: EHR – electronic health records

**Table 1** Methods of identifying Lyme disease patients with post-treatment Lyme disease syndrome (PTDLS)

| | Symptom diagnosis | PTLDS diagnosis | Text in encounter note |
|---|---|---|---|
| Symptom diagnosis ($n=638$), n, row % | | 8 (1.3) | 46 (7.2) |
| PTLDS diagnosis ($n=14$), n, row % | 8 (57.1) | | 10 (71.4) |
| Text in encounter note ($n=100$), n, row % | 46 (46.0) | 10 (10.0) | |

and linked these data to sequenced DNA from a large biorepository. We then examined potential shared associations with fibromyalgia and ME/CFS, common diagnoses with significant morbidity and symptoms similar to PTLDS. The GWAS identified two PTLDS loci that reached a suggestive significance threshold ($P<5\times10^{-7}$), but these were not associated with either fibromyalgia or ME/CFS.

Using an enhanced EHR-based algorithm for PTLDS, we classified 19.4% of individuals with Lyme disease with PTLDS, generally consistent with prior studies reporting PTLDS in 10 to 20% of individuals with Lyme disease [3]. Prior attempts to identify individuals with PTLDS have been limited to discrete data fields in the EHR and administrative claims data [16, 31]. This is particularly problematic as there are no billing codes for PTLDS specifically. We enhanced previously used approaches to identifying PTLDS, using physician-informed searches of clinician notes and the EDG diagnosis for PTLDS, available in EPIC EHR systems. A prior EHR-based study of Lyme disease stage found that among 18% of Lyme disease cases, there was information in the EHR regarding Lyme disease stage (e.g., early vs. disseminated Lyme disease) in free text notes that was not available in diagnoses [32]. In our study, discrete data on PTLDS-related symptoms captured less than half of the individuals who were described as having PTLDS in the clinician notes.

We hypothesized that the GWAS would identify SNPs that could be mapped to genes linked to potential pathways to PTLDS, including inflammatory or other immunologic response, and central sensitivity syndrome [2, 6]. Of the two loci that reached a suggestive significant threshold ($P<5\times10^{-7}$), neither had a clear functional role related to nearby genes. Our top index SNP, rs77857587, was in PTLDS with a long-range protein quantitative locus (pQTL) SNP, rs111774530, for the MARC2 protein. *MARC2* expression has been negatively associated with several immune checkpoints [29], immune regulators of both stimulatory and inhibitory pathways that regulate magnitude and duration of immune response [30]. Immunologic response has been posited to contribute to persistent symptoms after Lyme disease treatment [33]. Our results should be interpreted with caution as

**Table 2** Summary statistics for cases and controls across each phenotype included in the genetic analysis

| | PTLDS | | | | Fibromyalgia | | | | ME/CFS | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Variable** | **Cases** | | **Controls** | | **Cases** | | **Controls** | | **Cases** | | **Controls** | |
| Total (N/% total) | 670 | 20.16% | 2653 | 79.84% | 5,082 | 3.23% | 152,063 | 96.77% | 2,218 | 1.44% | 151,814 | 98.56% |
| Female (N/% strata) | 391 | 58.36% | 1,345 | 50.70% | 4,804 | 94.53% | 91,460 | 60.15% | 1,531 | 69.03 | 92,551 | 60.96% |
| Age (Mean/SD) | 54.21 | 14.79 | 55.71 | 14.85 | 55.68 | 13.48 | 57.05 | 17.11 | 57.34 | 15.88 | 56.98 | 17.03 |
| **Race/Ethnicity (N/% strata)** | | | | | | | | | | | | |
| White / European American | 652 | 97.31% | 2,611 | 98.42% | 4,717 | 92.82% | 142,962 | 94.01% | 2,118 | 95.49% | 142,602 | 93.93% |
| Black / African American | 2 | 0.30% | 12 | 0.45% | 107 | 2.11% | 3,143 | 2.07% | 29 | 1.31% | 3,182 | 2.10% |
| Hispanic / Latino | 15 | 2.24% | 18 | 0.68% | 202 | 3.97% | 4,126 | 2.71% | 42 | 1.89% | 4,208 | 2.77% |
| Asian | N/A | N/A | 9 | 0.34% | 15 | 0.30% | 723 | 0.48% | 9 | 0.41% | 719 | 0.47% |
| Other | 1 | 0.15% | 3 | 0.11% | 41 | 0.81% | 1,109 | 0.73% | 20 | 0.90% | 1,103 | 0.73% |

Abbreviations: PTLDS – post-treatment Lyme disease syndrome, ME/CFS - myalgic encephalomyelitis – chronic fatigue syndrome, SD – standard deviation

**Table 3** Summary of GWAS results

| Trait | Nearest Gene | MarkerID | CHR | POS (hg38) | ALT (EA) | REF (OA) | EAF | BETA | SE | *P* | *N* | *N*(Cases) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PTLDS | *IRX1* | rs77857587 | 5 | 3,343,244 | A | G | 0.0108 | 1.3050 | 0.2461 | 1.14E-07 | 3,322 | 670 |
| | *GAS2* | rs10833979 | 11 | 23,210,543 | T | C | 0.6113 | -0.3207 | 0.0627 | 3.19E-07 | 3,297 | 667 |
| PTLDS EUR | *IRX1* | rs62337498 | 5 | 3,356,556 | C | T | 0.0101 | 1.3657 | 0.2562 | 9.82E-08 | 3,258 | 652 |
| | *GAS2* | rs10833979 | 11 | 23,210,543 | T | C | 0.6115 | -0.3190 | 0.0634 | 4.87E-07 | 3,238 | 649 |
| PTLDS | *IRX1* | rs77857587 | 5 | 3,343,244 | A | G | 0.0103 | 1.3034 | 0.2597 | 5.19E-07 | 3,207 | 555 |
| (strict definition) | *GAS2* | rs10833979 | 11 | 23,210,543 | T | C | 0.0551 | -0.3144 | 0.0676 | 3.25E-06 | 3,182 | 552 |
| Fibromyalgia | *IRX1* | rs77857587 | 5 | 3,343,244 | A | G | 0.0108 | 0.1021 | 0.0946 | 0.2804 | 155,640 | 5,041 |
| | *GAS2* | rs10833979 | 11 | 23,210,543 | T | C | 0.6160 | -0.0041 | 0.0215 | 0.8492 | 155,640 | 5,041 |
| ME/CFS | *IRX1* | rs77857587 | 5 | 3,343,244 | A | G | 0.0109 | 0.1050 | 0.1434 | 0.4641 | 153,862 | 2,268 |
| | *GAS2* | rs10833979 | 11 | 23,210,543 | T | C | 0.6162 | 0.0153 | 0.0308 | 0.6203 | 153,862 | 2,268 |

Abbreviations: CHR - chromosome, POS (hg38) – position in Genome Reference Consortium Human Build 38, ALT - alternate allele, EA - effect allele, REF - reference allele, OA - other allele, EAF – effect allele frequency, SE - standard error, EUR - White/European American only analysis, PTLDS – post-treatment Lyme disease syndrome, ME/CFS - myalgic encephalomyelitis – chronic fatigue syndrome, *IRX1* - Iroquois Homeobox 1 gene, *GAS2* - Growth Arrest Specific 2 gene

our findings only met the suggestive significance threshold ($P<5\times10^{-7}$); however, MARC2 may warrant further consideration in future studies into the pathogenesis of PTLDS following Lyme infection.

In the absence of a replication cohort, we explored whether our PTLDS GWAS findings could be generalized in fibromyalgia or ME/CFS cohorts. We selected these conditions given their overlapping symptoms and at least one hypothesized common pathophysiology, CSS, central sensitization that is thought to involve hyperactivation of central neurons, leading to various synaptic and neurotransmitter changes [11]. Neither of the two suggestively significant ($P<5\times10^{-7}$) loci in the PTLDS analysis were associated with fibromyalgia or ME/CFS.

PTLDS is characterized by a range of symptoms, including pain, fatigue, and cognitive deficits [3]. This phenotypic heterogeneity may reflect a genetic heterogeneity. Uncovering the genetic pathways of such a condition generally requires large datasets. The lack of genome-wide significant findings in the PTLDS should be interpreted with caution as they may be due to insufficient sample size. Similarly, genetic factors for fibromyalgia and ME/CFS conditions have yet to be identified and studies have not been able to establish clear genetic association, potentially due to the high phenotypic heterogeneity of these conditions and a lack of high validity phenotyping algorithms for both related phenotypes [34, 35].

In addition to potential limitations of sample size, we did not require laboratory confirmation of Lyme disease to classify a patient as a Lyme disease case, thus there is some risk of misclassification. Per guidelines, laboratory measures are not required for Lyme disease diagnosis in all cases, specifically patients presenting early in the disease process with potential tick exposure and erythema migrans [36]. Requiring laboratory evidence would have excluded patients who were clinically diagnosed early in the disease process. We conducted this study in a region of the United States that is highly endemic to Lyme disease. The species of *Borrelia* that cause human infection differs in Europe and the United States and may explain some of the variation in the clinical manifestations of Lyme disease across countries [1]. Findings may not be generalizable to other regions of the world endemic to Lyme disease.

Another limitation of our study published is that our participants are predominantly White/European individuals (97%), limiting the generalizability of our findings to the general U.S. population. Having a diverse study population can introduce heterogeneity into an analysis and increase the likelihood of confounding due to population structure in GWAS. We chose to retain all eligible participants rather than restrict to one population group, but we present a sensitivity analysis in White/European

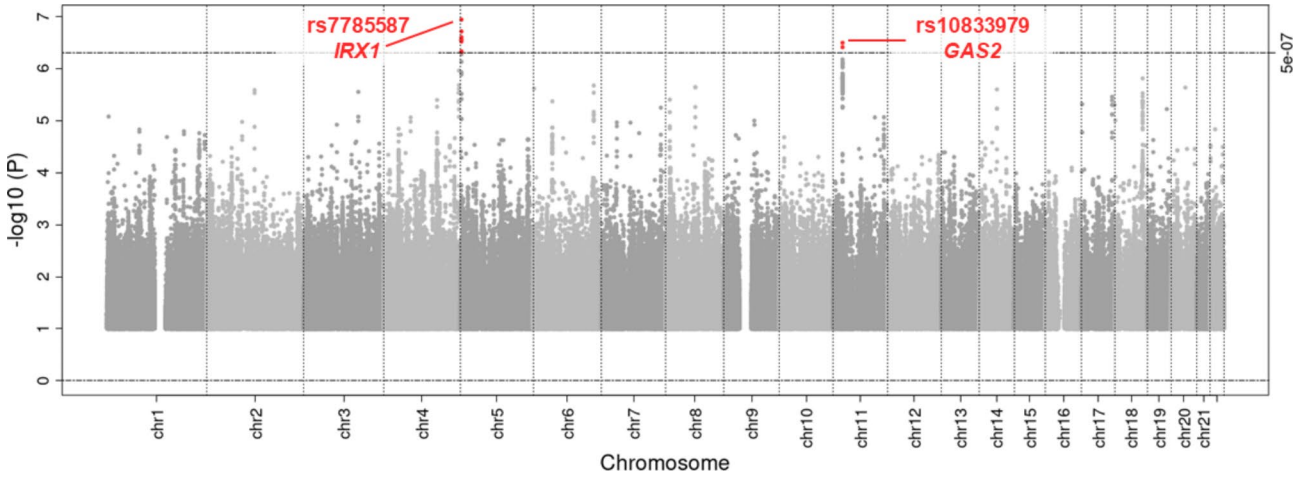


**Fig. 2** Manhattan plot of association results for GWAS (genome-wide association study) of PTLDS (post-treatment Lyme disease syndrome). Suggestively significant loci ($P < 5 \times 10^{-7}$) are annotated by the top index SNP and nearest protein coding gene

individuals to illustrate the robustness of our findings. Further, previous work has shown that including diverse participants while using appropriate analytical methods that account for population structure and phenotypic heterogeneity can enable new discoveries [33, 37] that generalize across populations. Thus, future work should include a more diverse study population to determine the generalizability of our findings across race/ethnic and ancestral groups.

There are a number of strengths to our study. We were able to identify PTLDS cases using both discrete and free-text EHR data. Free-text data has been found to contain information about Lyme disease not available in diagnosis information and has not previously been used to identify PTLDS [30]. By using a GWAS approach we were able to evaluate a range of potential genetic variants of interest, critical to studying a condition for which the pathogenesis remains poorly understood, and implicate a promising candidate gene for future interrogation (*MARC2*). Furthermore, we were able to evaluate the generalizability of our PTLDS GWAS findings in two separate cohorts of individuals with conditions that are common, can be disabling, and that have symptom profiles similar to PTLDS.

## Conclusion

In a region highly endemic to Lyme disease, we identified two PTLDS-associated loci that reached a suggestive significance ($P < 5 \times 10^{-7}$) threshold, one of which is linked with the MARC2 protein, a protein that has been related to multiple immune checkpoints. These findings were not observed in cohorts of individuals with fibromyalgia and ME/CFS. Further study is needed in a larger population to evaluate whether there is genetic evidence of the role of immune response in the occurrence of PTLDS.

### Abbreviations

| | |
|---|---|
| CSS | Central sensitization syndrome |
| CPT | Current Procedural Terminology |
| EDG | Electronic diagnosis group |
| EHR | Electronic health record |
| FUMA | Functional mapping and annotation analysis |
| GWAS | Genome-wide association study |
| GSA | Global Sequencing Array |
| HWE | Hardy-Weinberg equilibrium |
| IBD | Identity by descent |
| ICD | International Classification of Diseases |
| MARC | Mitochondrial Amidoxime Reducing Component protein |
| MAF | Minor allele frequency |
| ME/CFS | Myalgic encephalomyelitis – chronic fatigue syndrome |
| PTLDS | Post-treatment Lyme disorder |
| SNP | Single nucleotide polymorphism |
| TOPMed | Trans-Omics for Precision Medicine |

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12879-024-10238-x.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

### Author contributions

Conception: AGH, BSS; Design: AGH, BSS, AEJ; Data acquisition: AEJ, CMN, NSJ, AP; Data analysis: AEJ, CMN, NSJ, AP; Interpretation: AGH, AEJ, JA, AWR, BSS; Original draft of manuscript: AGH; Substantial contribution to revisions: AGH, BSS, AEJ, AWR, JA.

### Funding

This study was funded by the Commonwealth Universal Research Enhancement Program (PA Cure), Pennsylvania Department of Health. The funder did not have a role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.

**Data availability**
Availability of data and materials: The datasets analyzed during the current study are not publicly available to protect patient confidentiality, but may be available from the corresponding author on reasonable request.

## Declarations

**Ethics approval and consent to participate**
This study was approved by the Geisinger Institutional Review Board: 2019 – 0907. All participants provided written informed consent to participate in the MyCode project.

**Consent for publication**
Not applicable.

**Competing interests**
The authors declare no competing interests.

**Author details**
[1]Department of Population Health Sciences, Geisinger Health System, 100 N. Academy Avenue, Danville, PA 17822-4400, United States of America
[2]Division of Rheumatology, Department of Medicine, Lyme Disease Research Center, Johns Hopkins University School of Medicine, Baltimore, United States of America
[3]Department of Environmental Health and Engineering, Johns Hopkins University Bloomberg School of Public Health, Baltimore, United States of America

Received: 4 January 2024 / Accepted: 18 November 2024
Published online: 24 February 2025

**References**

1. Mead P. Epidemiology of Lyme disease. Infect Disease Clin. 2022;36(3):495–521.
2. Nemeth J, Bernasconi E, Heininger U, Abbas M, Nadal D, Strahm C, et al. Update of the Swiss guidelines on post-treatment Lyme disease syndrome. Swiss Med Wkly. 2016;146:w14353. https://doi.org/10.4414/smw.2016.14353
3. Marques A. Persistent symptoms after treatment of Lyme disease. Infect Disease Clin N Am. 2022;36:621–38.
4. Wormser GP, Dattwyler RJ, Shapiro ED, Halperin JJ, Steere AC, Klempner MS, et al. The clinical assessment, treatment, and prevention of Lyme disease, human granulocytic anaplasmosis, and babesiosis: clinical practice guidelines by the Infectious Diseases Society of America. Clin Infect Dis. 2007;45(7):941. https://doi.org/10.1086/508667
5. Aucott JN, Crowder LA, Kortte KB. Development of a foundation for a case definition of post-treatment Lyme disease syndrome. Int J Infect Dis. 2013;17(6):e443–9. https://doi.org/10.1016/j.ijid.2013.01.008
6. Aucott JN, Rebman AW, Crowder LA, Kortte KB. Post-treatment Lyme disease syndrome symptomatology and the impact on life functioning: is there something here? Qual Life Res. 2013;22(1):75–84. https://doi.org/10.1007/s11136-012-0126-6
7. Auwaerter PG, Bakken JS, Dattwyler RJ, Dumler JS, Halperin JJ, McSweegan E, et al. Antiscience and ethical concerns associated with advocacy of Lyme disease. Lancet Infect Dis. 2011;11(9):713–9. https://doi.org/10.1016/S1473-3099(11)70034-2
8. Dattwyler RJ, Wormser GP, Rush TJ, Finkel MF, Schoen RT, Grunwaldt E, et al. A comparison of two treatment regimens of ceftriaxone in late Lyme disease. Wien Klin Wochenschr. 2005;117(11–12):393–7.
9. Oksi J, Nikoskelainen J, Hiekkanen H, Lauhio A, Peltomaa M, Pitkäranta A, et al. Duration of antibiotic treatment in disseminated Lyme borreliosis: a double-blind, randomized, placebo-controlled, multicenter clinical study. Eur J Clin Microbiol Infect Dis. 2007;26(8):571–81. https://doi.org/10.1007/s10096-007-0340-2
10. Rebman AW, Aucott JN. Post-treatment Lyme disease as a model for persistent symptoms in Lyme disease. Front Med. 2020;7(57). https://doi.org/10.3389/fmed.2020.00057
11. Batheja S, Nields JA, Landa A, Fallon BA. Post-treatment Lyme syndrome and central sensitization. J Neuropsychiatry Clin Neurosci. 2013;25(3):176–86.
12. Bramwell KK, Ma Y, Weis JH, Chen X, Zachary JF, Teuscher C, et al. Lysosomal β-glucuronidase regulates Lyme and rheumatoid arthritis severity. J Clin Invest. 2014;124(1):311–20. https://doi.org/10.1172/JCI72339
13. Schröder NW, Diterich I, Zinke A, Eckert J, Draing C, von Baehr V, et al. Heterozygous Arg753Gln polymorphism of human TLR-2 impairs immune activation by Borrelia burgdorferi and protects from late stage Lyme disease. J Immunol. 2005;175(4):2534–40.
14. Strle K, Shin JJ, Glickstein LJ, Steere AC. Association of a toll-like receptor 1 polymorphism with heightened Th1 inflammatory responses and antibiotic-refractory Lyme arthritis. Arthritis Rheum. 2012;64(5):1497–507. https://doi.org/10.1002/art.34383
15. Carey DJ, Fetterolf SN, Davis D, Faucett WA, Kirchner HL, Mirshahi U. The Geisinger MyCode Community Health Initiative: an electronic health record-linked biobank for precision medicine research. Genet Med. 2016;18(9):906–13.
16. Moon KA, Pollak J, Hirsch AG, Aucott JN, Nordberg C, Heaney CD et al. Epidemiology of Lyme disease in Pennsylvania 2006–2014 using electronic health records. Tick and Tick-borne Diseases. 2018;Oct 26. pii: S1877-959X(18)30131-6. https://doi.org/10.1016/j.ttbdis.2018.10.010
17. Taliun D, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program. Biorxiv. 2019. https://doi.org/10.1101/563866
18. Das S, Forer L, Schönherr S, Sidore C, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48:1284–7.
19. Fuchsberger C, Abecasis GR, Hinds D. A. minimac2: faster genotype imputation. Bioinformatics. 2014;31(5):782–4.
20. Lin DY, Tao R, Kalsbeek W, Zeng D, Gonzalez F, Fernández-Rhodes L, Graff M, Koch G, North KE, Heiss G. Genetic association analysis under complex survey sampling: the hispanic community health study/study of Latinos. Am J Hum Genet. 2014;95(6):675–88.
21. Staples J, Qiao D, Cho MH, Silverman EK, Nickerson DA, Below JE, University of Washington Center for Mendelian Genomics. PRIMUS: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. Am J Hum Genet. 2014;95:553–64.
22. Staples J, Maxwell EK, Gosalia N, Gonzaga-Jauregui C, Snyder C, Hawes A, Penn J, Ulloa R, Bai X, Lopez AE, Van Hout CV, O'Dushlaine C, Teslovich TM, McCarthy SE, Balasubramanian S, Kirchner HL, Leader JB, Murray MF, Ledbetter DH, Shuldiner AR, Yancoupolos GD, Dewey FE, Carey DJ, Overton JD, Baras A, Habegger L, Reid JG. Profiling and leveraging relatedness in a precision medicine cohort of 92,455 exomes. Am J Hum Genet. 2018;102(5):874–89. https://doi.org/10.1016/j.ajhg.2018.03.012
23. Winkler TW, et al. Quality control and conduct of genome-wide association meta-analyses. Nat Protoc. 2014;9:1192–212.
24. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8(1):1826.
25. Watanabe K, Mirkov MU, de Leeuw CA, van den Heuvel MP, Posthuma D. Genetic mapping of cell type specificity for complex traits. Nat Commun. 2019;10(1):3222. https://doi.org/10.1038/s41467-019-11181-1
26. Kircher M, et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310–5.
27. Boyle AP, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22:1790–7.
28. He B, Shi J, Wang X, Jiang H, Zhu H. Genome-wide pQTL analysis of protein expression regulatory networks in the human liver. BMC Biol. 2020;18(1):97. https://doi.org/10.1186/s12915-020-00830-3
29. Wu D, Liang S, Guo H, Zhang S, Yang G, Yuan Y, Liu L. Downregulation of MARC2 promotes immune escape and is associated with immunosuppression of hepatocellular carcinoma. Front Genet. 2022. https://doi.org/10.3389/fgene.2021.790093
30. Hu F, Liu C, Liu L, Zhang Q, Guo A. Expression profile of immune checkpoint genes and their roles in predicting immunotherapy response. Brief Bioinform. 2021;22(3):1–12.
31. Chung MK, Caboni M, Stranditz P, D'Onofrio A, Patel CJ. Systematic comparisons between Lyme disease and post-treatment Lyme disease syndrome in the U.S. with administrative claims data. eBioMedicine. 2023;90. https://doi.org/10.1016/j.ebiom.2023.104524
32. Moon K, Pollak J, Poulsen M, Heaney C, Hirsch AG, Schwartz B. Risk factors for Lyme disease stage and manifestation using electronic health records. BMC Infect Disease. 2022:1–13. https://doi.org/10.1186/s12879-021-06959-y
33. Aucott JN, Soloski MJ, Rebman AW, Crowder LA, Lahey LJ, Wagner CA, Robinson WH, Bechtold KT. CCL19 as a chemokine risk factor for posttreatment

Lyme disease syndrome: a prospective clinical cohort study. Clin Vaccine Immunol. 2016;23(9):757–66.
34. Docampo E, Escaramis G, Gratacos M, et al. Genome-wide analysis of single nucleotide polymorphisms and copy number variants in fibromyalgia suggest a role for the central nervous system. Pain. 2014;155:1102–9.
35. Das S, Taylor K, Sardell J, Gardner S. Genetic risk factors for ME/CFS identified using combinatorial analysis. J Translational Med. 2020;20:598.
36. Lantos PM, Rumbaugh J, Bockenstedt LK, Falck-Ytter YT, Auero-Rosenfeld ME, Auwaerter PG, et al. Clinical practice guidelines by the Infectious Diseases Society of America (IDSA), American Academy of Neurology (AAN), and American College of Rheumatology (ACR): 2020 guidelines for the prevention, diagnosis and treatment of Lyme disease. Clin Infect Dis. 2021;72(1):e1–48.
37. Zhang BC, Biddanda A, Freyr Gunnarsson A, Cooper F, Francesco Palamara P. Biobank-scale inference of ancestral recombination graphs enables genealogical analysis of complex traits. Nat Genet. 2023;55:768–76.

## Publisher's note