medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Immunosequencing of the T-cell receptor repertoire reveals signatures specific for diagnosis and characterization of early Lyme disease

Julia Greissl*[a], Mitch Pesesky*[b], Sudeb C. Dalai[b,c], Alison W. Rebman[d], Mark J. Soloski[d], Elizabeth J. Horn[e], Jennifer N. Dines[b], Rachel M. Gittelman[b], Thomas M. Snyder[b], Ryan O. Emerson[b], Edward Meeds[a], Thomas Manley[b], Ian M. Kaplan[b], Lance Baldo[b], Jonathan M. Carlson[a,†], Harlan S. Robins[†b], John N. Aucott[†d]

**Julia Greissl and Mitch Pesesky contributed equally to this study.*
[†] Jonathan M. Carlson, Harlan S. Robins, and John N. Aucott contributed equally to the oversight of this study.

[a]Microsoft Research, Redmond, Washington, USA, and Cambridge, UK

[b]Adaptive Biotechnologies, Seattle, Washington, USA

[c]Stanford University School of Medicine, Stanford, California, USA

[d]Lyme Disease Research Center, Division of Rheumatology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[e]Lyme Disease Biobank, Portland, Oregon, USA

Corresponding author:
John N. Aucott, MD
Lyme Disease Research Center
The Johns Hopkins University
2360 W. Joppa Road
Lutherville, MD, USA 21093
Tel: 410-616-7596
Fax: 410-367-2371
Email: jaucott2@jhmi.edu

Key words: Lyme disease, *Borrelia burgdorferi*, T-cell receptor, diagnostic, next-generation sequencing

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

ABSTRACT

Lyme disease, the most common tick-borne illness in the United States, is most frequently caused by infection with *Borrelia burgdorferi*. Although early antibiotic treatment can prevent development of severe illness and late manifestations, diagnosis is challenging in patients who do not present with a typical erythema migrans rash. To support a diagnosis of Lyme disease in such cases, guidelines recommend 2-tiered serologic testing. However, 2-tiered testing has numerous limitations, including ambiguity in interpretation and lower sensitivity in early disease. We developed a diagnostic approach for Lyme disease based on the T-cell response to *B. burgdorferi* infection by immunosequencing T-cell receptor (TCR) repertoires in blood samples from 3 independent cohorts of patients with laboratory-confirmed or clinically diagnosed early Lyme disease, as well as endemic and non-endemic controls. We identified 251 public, Lyme-associated TCRs that were used to train a classifier for detection of early Lyme disease with 99% specificity. In a validation cohort of individuals with early Lyme disease, TCR testing demonstrated a 1.9-fold increase in sensitivity compared to standard 2-tiered testing (STTT; 56% versus 30%), with a 3.1-fold increase ≤4 days from the onset of symptoms (44% versus 14%). TCR positivity predicted subsequent seroconversion in 37% of initially STTT-negative patients, suggesting that the T-cell response is detectable before the humoral response. While positivity for both tests declined after treatment, greater declines in posttreatment sensitivity were observed for STTT compared to TCR testing. Higher TCR scores were associated with clinical measures of disease severity, including abnormal liver function test results, disseminated rash, and number of symptoms. A subset of Lyme-associated TCRs mapped to *B. burgdorferi* antigens, demonstrating high specificity of a TCR immunosequencing

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

approach. These results support the clinical utility of T-cell–based testing as a sensitive and specific diagnostic for early Lyme disease, particularly in the initial days of illness.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

INTRODUCTION

Lyme disease, the most common tick-borne illness in the United States (U.S.), has an estimated incidence of >450,000 new cases annually (1–3). In the U.S., Lyme disease is caused by infection with the spirochetal bacterium *Borrelia burgdorferi* (or rarely *B. mayonii*) transmitted from infected *Ixodes* ticks (4–6). Lyme disease is also among the most widespread tick-borne diseases worldwide, although in Eurasia, *B. afzelii* and *B. garinii* (in addition to *B. burgdorferi*) commonly cause infection (1, 6).

In the days to weeks after the initial tick bite, early symptoms may include a characteristic erythema migrans (EM) rash and nonspecific flu-like symptoms. Individuals presenting at later stages may exhibit signs and symptoms of disseminated infection affecting the joints, nervous system, or heart (7). The clinical manifestations of disseminated infection vary based on the infecting *Borrelia* species and region, with joint-related symptoms being more common in North America and severe neurological and chronic skin manifestations occurring more frequently in Eurasia (1, 8–10). Potentially debilitating late manifestations include arthritis, encephalopathy, encephalomyelitis, peripheral neuropathy, or acrodermatitis chronica atrophicans. Early diagnosis and treatment of Lyme disease has been shown to prevent severe illness and the development of late objective manifestations of disease (4, 11).

Given the high clinical index of suspicion of EM rash for Lyme disease and the poor sensitivity and specificity of currently available diagnostic assays, patients with EM rash can be treated immediately without further testing (4). However, while the CDC reports that the majority of patients develop an EM rash, Lyme-associated rashes can be mistaken for other conditions or

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

may go unnoticed or unreported due to their location, highly variable appearance, and/or transient nature (12–15). The Infectious Diseases Society of America (IDSA) recommends serologic testing to support a diagnosis of Lyme disease for patients with atypical EM rash, as well as for patients in appropriate epidemiologic settings who have symptoms and exposure compatible with disseminated infection, such as arthritis, central nervous system involvement, or acute myocarditis/pericarditis (16). Some studies have reported that EM rash is absent in 50% to 60% of laboratory-confirmed cases of Lyme (17, 18), highlighting the difficulty of reaching a definitive clinical diagnosis.

The primary CDC-recommended testing option for Lyme disease is standard 2-tiered testing (STTT), which combines an enzyme-linked immunosorbent assay (ELISA) with a subsequent, more specific immunoblot assay to detect IgM and IgG antibodies against *B. burgdorferi* for positive or equivocal samples (19, 20). Even in individuals with rigorously defined clinical Lyme disease, STTT exhibits poor sensitivity (25%–50%) during the acute phase of infection as antibody responses are developing (21), and misinterpretation of weak IgM immunoblot bands can lead to inter- and intra-laboratory variability (22). However, in untreated individuals with advanced-stage disease (months to years after initial infection, when serious late-stage complications may have developed), the sensitivity of STTT may approach ~90% (23).

Real-world evidence suggests that available testing modalities are not meeting the needs of current clinical practice. Physician surveys highlight confusion about the interpretation of results for different immunoglobulin isotypes (24) and demonstrate that the majority of serologic tests are requested in the early stages of infection, when sensitivity of 2-tiered testing is lower (25). Indeed, up to 60% of patients testing negative early in the course of infection may

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

have Lyme disease with early seronegativity (21, 23, 26–31). Diagnostic interpretation is also complicated by the predictive value of serology testing, which is much lower in non-endemic regions where the pre-test probability of Lyme disease is low (20, 32). False-negative results have been reported to occur in patients with early antibiotic treatment or concurrent bacterial infections that can decrease antibody responses to *Borrelia* or block IgG seroconversion (32, 33). A small percentage of patients with Lyme disease may also test seronegative due to development of a cellular response in the absence of a humoral response (32). Furthermore, because serology assays cannot distinguish between active and past infections, a subset of the 5% to 10% of symptomatic individuals in endemic areas who test positive by STTT may in fact have an unrelated illness (24, 32, 34). These data underscore a number of unmet clinical needs for novel methods that can facilitate more sensitive and specific diagnosis of Lyme disease, especially in the early stages of infection.

Diagnostic tests based on the cellular immune response can address some of the limitations of serology-based testing, as infection with *B. burgdorferi* has been shown to elicit a T-cell response that may exhibit different kinetics than the humoral response (35, 36). Evaluation of cytokine/chemokine profiles suggests that an active T-cell response is induced during the acute phase of infection, even in the absence of seroconversion, and returns to normal levels after treatment and symptom resolution (37). In individuals with persistent symptoms, Th1 responses in affected tissues have been linked to pathogenic inflammation (38–40). In contrast, humoral responses vary widely, with some cases demonstrating attenuated responses and lack of IgM to IgG seroconversion, and other cases demonstrating antibody persistence for decades

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(32, 33, 41). These data suggest that assays interrogating the T-cell response may have utility for aiding in the diagnosis of Lyme disease during early illness, as well as late manifestations.

High-throughput sequencing of the T-cell receptor (TCR) repertoire can be used to identify disease-specific TCR sequences expressed by T-cell clones that have undergone antigen-driven expansion and persist in the memory compartment. While the diversity of TCR recombination means that most TCR responses are "private" and infrequently observed in other individuals, part of the T-cell response to a disease is "public," with identical amino acid sequences observed across multiple individuals, particularly those with shared HLA backgrounds (42). Such disease-associated TCRs can be identified using a case/control design, as previously described for cytomegalovirus (CMV) (43) and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (44) viral infections, and matched to specific antigens through multiplex identification of antigen-specific T-cell receptors (MIRA) (43). Because these public clones are antigen- and HLA-specific, they serve as a signature of infection in a given HLA context (43, 44).

We have previously shown that classifiers based on quantification of disease-associated public TCR sequences can be used for sensitive identification of past infection with CMV (43) or SARS-CoV-2 (44). Such classifiers leverage the relative frequency of disease-associated sequences within the repertoire, measures that have been shown to be associated with disease severity in the setting of SARS-CoV-2 (45, 46). Clinical validation of a T-cell assay for SARS-CoV-2 utilizing a similar methodology demonstrated high positive percent agreement (>94.5%) and negative percent agreement (~100%) with reverse transcriptase PCR (RT-PCR) for detection of past SARS-CoV-2 infection (46, 47). However, this approach has not previously been applied to bacterial disease.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

In the present study, we describe an approach for measuring the T-cell adaptive immune response in early Lyme disease using TCRβ sequencing from blood samples. We analyzed blood samples from three independent cohorts of patients with laboratory-confirmed and/or clinically diagnosed early Lyme disease acquired from the Lyme Disease Biobank (LDB) of the Bay Area Lyme Foundation; Boca Biolistics (Pompano Beach, FL); and the Study of Lyme disease Immunology and Clinical Events (SLICE) cohort collected at the Johns Hopkins University (JHU) Lyme Disease Research Center. We first identified samples from STTT-positive patients with signs and symptoms of early Lyme disease from the LDB and Boca Biolistics cohorts and combined them with controls from a database of healthy individuals to create a primary training dataset to identify Lyme-associated, public TCR signatures. We then validated the diagnostic performance of the resulting classifier using samples from patients with clinically diagnosed Lyme disease from the JHU cohort as well as endemic controls held out from the LDB and JHU cohorts and a database of individuals presumed negative for Lyme disease. Finally, we assessed the correlation between T-cell responses and clinical features of Lyme disease and mapped a subset of the identified Lyme-associated TCRs to specific *B. burgdorferi* antigens, demonstrating the biologic specificity of TCR immunosequencing.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

RESULTS

Identification of shared Lyme disease-associated TCRs across the population

To identify public TCRs associated with early Lyme disease, we designed a case/control training dataset consisting of patients identified from the LDB and Boca Biolistics cohorts who presented with STTT-positive early Lyme disease prior to 2019 (*n* = 72) and control repertoires (*n* = 2981) from a database of healthy individuals from endemic and non-endemic regions recruited for other studies and presumed to be negative for early Lyme disease (Fig. 1A). Table S1 summarizes the cohorts who provided samples used in this study. Public, Lyme disease-associated TCRs, referred to as "enhanced sequences," were identified primarily based on statistical enrichment in cases, as described in the Methods. Overall, we identified 251 enhanced sequences associated with early Lyme disease.

Enhanced TCR sequences are highly specific for identifying early Lyme disease

As previously observed in viral infections (43, 44), comparison of the numbers of enhanced sequences and total unique productive TCR rearrangements among case and control samples suggests that the total number of disease-associated enhanced sequences in a repertoire is a highly specific biomarker for Lyme disease (Fig. 1A). To leverage this biomarker as a diagnostic classifier, we modeled the number of enhanced sequences as a logistic-growth function of the number of unique productive TCRs sampled from a repertoire and fit this model to the 2981 control repertoires in the training data (Fig. 1A; see black line representing the model fit in Fig.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

1A). The resulting model compares the number of observed versus expected enhanced sequences in a repertoire, given the number of observed unique TCRs, quantified as the number of standard deviations from the expected value (red dashed lines, Fig. 1A). This approach carefully controls specificity by considering thousands of control repertoires. The final positive/negative call threshold was set to a specificity of 99% on an independent set of endemic control samples (*n* = 2627) (Fig. 1B; Table S1). To confirm the specificity and generalizability of the classifier and call threshold, we applied the resulting model to a holdout set of samples from the LDB cohort collected in 2019 that included both laboratory-confirmed positive (by STTT, PCR, and/or culture) cases of early Lyme and laboratory-confirmed negative (by STTT) endemic controls with no history of Lyme or tick-borne infection (Fig. 1C). Overall, 8 of 15 (53%) early Lyme disease samples and 0 of 48 (0%) endemic control samples were identified as TCR-positive by the classifier.

TCR repertoire analysis is more sensitive than STTT for identifying early Lyme disease and frequently precedes STTT seroconversion

To further evaluate the performance of the TCR assay, we validated the TCR classifier using samples from STTT-positive and STTT-negative patients with clinically diagnosed early Lyme disease enrolled in the JHU cohort. Application of the TCR classifier revealed that median TCR model scores for patients with early Lyme disease were higher than those of individuals defined as endemic controls based on cohort-specific criteria (Fig. 2A; Table S1). Overall, 118 of 211 (56%) patients diagnosed with early Lyme disease were classified as TCR-positive (Table 1). By

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

comparison, 64 of 211 (30%) patients were STTT-positive, indicating that use of the TCR assay nearly doubled the number of clinical Lyme disease cases identified as positive (1.9-fold increase in sensitivity). Only 32 of 2631 (1.2%) endemic control samples tested TCR-positive (0 of 115 in LDB; 1 of 45 in JHU; and 31 of 2471 repertoires from our database from individuals with unknown Lyme disease status living in Lyme-endemic regions in the US and Europe; Fig. 2A). Of note, PCR testing results for blood and/or skin biopsy were available for a subgroup of 57 individuals in the JHU cohort; of those, 12 individuals tested negative by both STTT and PCR, raising the possibility that these individuals had a non-Lyme tick-borne illness, such as Southern tick-associated rash illness (STARI) (48). The majority (11/12) of these individuals were TCR-negative; excluding them from the analysis did not appreciably alter performance characteristics (sensitivities of 59% versus 32% for TCR assay and STTT, respectively).

The sensitivity of both TCR testing and STTT were lower in early illness and increased with days since symptom onset (Fig. 2B). While the sensitivity of TCR testing was greater than STTT at all periods of illness evaluated, the greatest performance advantage of TCR testing was observed within the first week after symptom onset, when STTT sensitivity was below 30% (≤4 days: 44% versus 14%; 5 to 8 days: 57% versus 29%; >8 days: 68% versus 51%, for TCR and STTT testing, respectively). These data indicate that TCR sequencing can identify early Lyme disease with significantly greater sensitivity than standard antibody-based testing, particularly in the initial days of acute illness, while also maintaining high specificity.

We next compared the agreement of TCR and STTT results, showing that TCR testing was positive in 59 of 64 (92%) STTT-positive cases (58 of 61 [95%] STTT-positive by IgM), as well as 59 of 147 (40%) STTT-negative cases (Table 1). Of the 59 TCR-positive/STTT-negative individuals,

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

22 (37%) subsequently seroconverted between study enrollment and the first posttreatment follow-up visit (~3 weeks after enrollment), while only 16 of 88 (18%) individuals who were TCR-negative at baseline seroconverted over the same time period (*P*=0.01, Fisher's exact test). Stratification of the JHU cohort by initially STTT-positive, posttreatment seroconverter, or persistent STTT-negative demonstrated that median TCR model scores (Fig. 2C) and classifier sensitivity (Fig. 2D, Table 1) were highest among individuals who were STTT-positive at enrollment, intermediate among those who seroconverted posttreatment, and lowest among individuals who remained persistently STTT-negative. Taken together, these data indicate that while the presence of a detectable T-cell response is strongly correlated with a detectable antibody response, earlier maturation of the T-cell response may allow for enhanced sensitivity of a TCR-based diagnostic during early phases of *B. burgdorferi* infection.

Disease-associated TCRs and seropositivity wane after treatment

Previous data suggest that the dynamics of T-cell and humoral immune responses differ in *B. burgdorferi* infection (36). To better understand the dynamics of the T-cell response posttreatment, we evaluated TCR repertoires in longitudinal samples from individuals enrolled in the JHU cohort. Patients initiated 3 weeks of oral doxycycline treatment within ±72 hours of enrollment, with samples collected at enrollment, immediately after treatment (~3 weeks after enrollment), and 6 months posttreatment. Immunosequencing of samples collected during these timepoints (*n* = 161 patients with available samples at all timepoints) revealed that TCR responses waned significantly in the 6 months following treatment (Fig. 3), differing from our

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

previous observations characterizing the T-cell response in patients infected with SARS-CoV-2 (46). Median model scores decreased from 6.1 to 2.5, and model sensitivity decreased from 56% (91/161) at enrollment to 32% (51/161) 6 months posttreatment. Notably, the sensitivity of STTT also declined over the same time period, from 33% at enrollment to 12% at 6 months posttreatment (14 of 115 patients with available STTT results), consistent with previous reports indicating that IgG seroconversion is often absent among IgM-positive individuals treated early in infection (33). Similar to the results shown in Fig. 2C, we observed that TCR model scores were higher across all timepoints among individuals who at baseline were STTT-positive compared those who were STTT-negative (Fig. 3).

T-cell responses correlate with clinical measures of Lyme disease severity

The strong correlation observed between antibody and T-cell responses highlights the interconnectedness of the immune response in early Lyme disease, and may also reflect underlying pathogen burden, disease severity, or other clinical measures that drive the immune response. We therefore explored potential associations between clinical parameters previously reported in the JHU study (49) and the strength of the T-cell response as measured by the TCR model score at diagnosis. In both univariate analyses (Fig. 4) and a multiple regression model (Table S2) that adjusted for sex, age, and serostatus, higher TCRs scores were associated with markers of disease severity, including elevated liver function tests, disseminated rash, and the number of Lyme disease-associated symptoms. Sex, age, size of rash, and lymphocyte count were not associated with a difference in TCR model scores in this cohort (Table S2).

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Defining the antigen specificity of Lyme-associated TCRs

To evaluate the potential breadth of antigens detected by TCRs included in our Lyme classifier, we clustered enhanced sequences by sequence similarity, as described in the Methods. We identified 6 clusters of at least 5 sequences each, which together accounted for 105 of the 251 (42%) enhanced sequences (Table 2). Notably, in 5 of 6 clusters, statistical assignment of individual TCRs to HLA subtypes resulted in a consistent HLA assignment for the cluster, supporting the conclusion that clustered TCRs react to the same antigen and providing a putative HLA restriction for that antigen (Table 2). All assigned HLA subtypes were class II heterodimers, consistent with the prediction that T-cell responses to bacterial antigens will be predominantly HLA-II–restricted CD4+ T cells.

These analyses suggest that >40% of the enhanced TCR sequences included in our Lyme classifier recognize one of 6 specific HLA-restricted peptides. To further characterize the antigen specificity of Lyme-associated TCRs, we used MIRA to identify target TCR epitopes. We first synthesized 777 query peptides derived from 26 *B. burgdorferi* proteins and assigned either individual peptides or groups of related peptides to one of 426 unique MIRA pools, or "addresses," as described in the Methods. MIRA was then performed on T cells derived from peripheral blood mononuclear cells (PBMCs) collected from 395 healthy individuals using a version of the assay that selects for HLA-II–restricted CD4+ T cells.

One cluster (Table 2, cluster 6) contained 6 TCR sequences that all mapped to the same antigen by MIRA (MIINHNTSAINASRNNG from the *B. burgdorferi* flagellin B (FlaB) protein,

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

WP_002661938.1). Each of these 6 sequences was found in at least 3 individuals (range, 3 to 25) in the MIRA analysis, and at least 1 of these sequences was found in 26 individuals. All 25 of the 26 individuals who were HLA-typed expressed *HLA-DRB3*02:02*, the HLA associated with cluster 6. By comparison, the background expression frequency of *HLA-DRB3*02:02* was 14% among 394 HLA-typed individuals assessed by MIRA. Selective recognition of a *B. burgdorferi* antigen by these TCRs is also consistent with immunosequencing results showing that they were observed in 21% of JHU cases, but only 5% of holdout endemic controls (Fig. 5, FlaB (A)). Sensitive detection of an immune response targeting FlaB differentiates TCR testing from STTT, as antibodies to the FlaB protein used for immunoblotting in STTT are known to have low specificity (50).

By applying MIRA-based antigen assignment to enhanced sequences that were not clustered (but that were present in ≥2 individuals assessed by MIRA), we were able to assign antigens for 3 additional TCRs (Table S3). One TCR mapped to the same FlaB antigen as cluster 6 and was observed in MIRA experiments from 15 individuals. This sequence also had the same V gene, J gene, CDR3 length, and HLA association as the members of cluster 6, but did not meet our conservative clustering threshold due to a difference of 2 amino acids in the CDR3 region. Another enhanced sequence mapped to a different antigen from FlaB (SSGYRINRASDDAAGMG) and was found in 3 individuals by MIRA. This enhanced sequence was also associated with *HLA-DRB3*02:02* in the JHU cohort, and all 3 individuals from the MIRA analysis expressed this HLA. The third MIRA-assigned enhanced sequence mapped to DbpA in 4 individuals and could not be assigned an HLA based on our data. Each of these enhanced sequences was present in a higher proportion of JHU cases than endemic controls (Fig. 5).

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Finally, to evaluate the specificity of our approach for mapping Lyme disease-specific TCRs to specific *B. burgdorferi* antigens, we compared the Lyme-associated enhanced sequences to a set of TCRs from 507 individuals that were previously mapped to 325 SARS-CoV-2 antigen pools by MIRA (44, 51), once again limiting the analysis to TCRs identified in ≥2 distinct individuals. We identified no matches between SARS-CoV-2–associated TCRs and Lyme-associated enhanced sequences. Collectively, these results demonstrate the functional relevance of the TCRs included in the Lyme classifier and confirm the specificity of our approach for identifying Lyme disease.

DISCUSSION

We describe an approach for diagnosis of early Lyme disease from blood samples based on high-throughput TCR sequencing. Identification of 251 Lyme-associated enhanced TCR sequences served as the basis for training a classifier capable of sensitive and specific detection of Lyme disease across 3 independent cohorts of patients with laboratory-confirmed and/or clinically diagnosed early Lyme disease (LDB, Boca, and JHU). Validation of the classifier demonstrated that the T-cell assay identifies patients with early Lyme disease with a 1.9-fold improvement in sensitivity compared to STTT (56% versus 30%), while maintaining a specificity of 99%. Enhanced sensitivity was most apparent in early illness (44% versus 14%, or 3.1-fold increase in sensitivity ≤4 days since symptom onset), and TCR positivity was predictive of subsequent STTT seroconversion in 37% of initially STTT-negative individuals. T-cell testing was also more sensitive than STTT for identification of Lyme disease posttreatment (32% versus 12%

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

at 6 months posttreatment), including in patients who did not undergo IgG seroconversion between acute and convalescent samples. Higher TCR scores were associated with clinical measures of disease, including elevated liver function tests, disseminated rash, and number of disease-associated symptoms. Finally, we demonstrate that a subset of the identified Lyme-associated TCRs map to known *B. burgdorferi* antigens, supporting the high biologic specificity of a TCR immunosequencing approach.

Results from this study highlight the potential utility of a T-cell–based diagnostic for identification of Lyme disease during early stages of infection. A recent review comparing the sensitivity of 2-tiered testing algorithms in patients with early Lyme disease reported ranges of 25% to 50% for STTT (21), which together with our data, suggest that the sensitivity of a TCR-based diagnostic is greater than that of STTT in early-stage disease. In addition, the ability of the TCR assay to identify Lyme disease in a large proportion of STTT-negative individuals prior to seroconversion indicates that the TCR response may be detectable before the humoral response by any serologic testing modality. These data imply that T-cell activation precedes and may be required for some aspects of the humoral response, although both T-cell–dependent and independent responses have been implicated in clearance of *Borrelia* infection (52, 53). Given the high prevalence of testing performed in patients during the early stages of infection, when sensitivity of 2-tiered testing is poor (25), as well as absence of a detectable serological response in almost half of individuals with PCR-confirmed *B. burgdorferi* infection (48), an alternative approach that measures a different aspect of the immune response provides important clinical utility. Furthermore, a TCR-based assay with a clear threshold for positivity may alleviate confusion around the clinical interpretation of serology results (24).

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

In addition to establishing an unambiguous and easily interpretable cutoff for positivity, the TCR assay score may also serve as a semi-quantitative proxy for disease activity. Correlation of TCR scores with clinical measures of disease, such as number of reported symptoms and disseminated rash, suggests that the magnitude of the T-cell response is associated with the degree of symptomatology. Furthermore, longitudinal analyses show that the TCR score decreases with time posttreatment, consistent with diminishment of the T-cell response with resolution of disease. However, while TCR positivity was associated with STTT positivity at enrollment (92%), the T-cell response did not decline as rapidly as serologic responses following treatment. This observation, as well as the increased sensitivity of the assay over STTT, suggests that TCR testing may be able to identify Lyme disease even in the absence of acute (IgM) to convalescent (IgG) seroconversion, a common occurrence among individuals treated early in the course of disease (33), and supports the role of T-cell–based testing as both an alternative and complementary method for diagnosis. Importantly, 11 of 12 individuals who presented with EM rash, but were both STTT-negative and *B. burgdorferi* PCR-negative, were also TCR-negative, calling into question the diagnosis of Lyme disease in these individuals and suggesting that T-cell–based testing may aid in the differential diagnosis of Lyme disease and similar tick-borne illnesses with overlapping manifestations, such as STARI (48).

Further studies are needed to understand the utility of TCR testing in patients with later stages of Lyme disease and long-term sequelae. Approximately 10% to 20% of patients treated for Lyme disease experience long-term symptoms lasting ≥6 months after treatment, known as posttreatment Lyme disease syndrome (PTLDS) (4, 54). Elevated IL-23 and CCL19 levels in individuals with PTLDS compared to those with symptom resolution suggest a role for

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

persistent T-cell responses in ongoing disease (55, 56). Consistent with this hypothesis, analysis of cerebrospinal fluid and synovial fluid from patients with Lyme neuroborreliosis or late Lyme arthritis has revealed the presence of Th1 cells that may contribute to pathogenic inflammation (38–40). Unlike antibody levels, which in some cases may remain elevated for decades in the absence of symptoms (41), the mechanistic link between Lyme pathology and aberrant T-cell responses may provide a means for diagnosis of late disease, while minimizing the potential for a misleading positive result caused by a previous exposure unrelated to the current symptomatology (37, 41). Given that molecular mimicry of tissue autoantigens has been proposed as a potential mechanism underlying late manifestations of Lyme disease (38), characterization of immunodominant epitopes during early infection may also provide insight into late sequelae and persistent symptoms. By combining MIRA and sequence-based analysis, we were able to identify antigens recognized by a subset of public Lyme-associated TCRs identified in this study. Future studies comparing whether enhanced TCR sequences associated with specific epitopes are preferentially associated with early versus late Lyme disease may provide insights into the pathophysiology of disease, as well as a means to predict which patients are most likely to develop late manifestations.

Application of the present TCR classifier as a diagnostic assay (T-Detect™ Lyme) will be further evaluated in order to support its clinical utility relative to 2-tiered serologic testing. While the present analysis is limited to samples previously collected from well-defined prospective cohorts of clinically confirmed and/or laboratory-confirmed early Lyme disease, additional prospective clinical validation studies are needed to further characterize the advantages of TCR testing relative to serology in scenarios where the spectrum of presenting illness may vary. In

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

addition, evaluation of potential assay cross-reactivity against other pathogens is also needed, though antigen mapping indicates that a subset of the identified TCRs are highly-specific for known *B. burgdorferi* proteins.

Results of this study demonstrate that TCR testing can have high clinical utility as a sensitive and specific diagnostic for Lyme disease. Diagnostic validation of our TCR classifier indicates that analysis of the T-cell response may facilitate diagnosis of early Lyme disease prior to detection of the humoral response, allowing for earlier recognition of disease and initiation of antimicrobial treatment to prevent the development of more severe illness in patients who lack definitive clinical signs/symptoms. Although this is the first study to evaluate the utility of TCR immunosequencing to identify acute bacterial infection, the observed diagnostic performance is consistent with that of similar TCR-based classifiers for identification of past CMV or SARS-CoV-2 infection (43, 44, 46, 47). These studies all leveraged a standardized approach to immunosequencing of the TCR repertoire from blood samples, followed by application of an algorithm designed to yield clear positivity thresholds for identification of disease cases. In addition, because the algorithms are based on statistical association of TCRs with disease, diagnostic sensitivity is expected to improve as the size of training data increases, while collection of large numbers of case/control samples across a range of environmental contexts will allow for detailed characterization and further generalization of the classifiers. Collectively, these studies indicate that characterization of the adaptive immune response through sequence-based identification of public, disease-specific TCRs is a powerful and generalizable approach to aid in diagnosing disease.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

ACKNOWLEDGMENTS

We are grateful to the research participants who contributed samples and data used in this study and to the physicians/health care providers who facilitated recruitment. Medical writing and editorial support were provided by Melanie Styers of BluPrint Oncology Concepts and Kristin MacIntosh of Adaptive Biotechnologies.

M.P., J.N.D., R.M.G., T.M.S., and I.M.K. declare employment and equity ownership in Adaptive Biotechnologies. S.C.D. declares employment with Adaptive Biotechnologies and Stanford University School of Medicine, and equity ownership in Adaptive Biotechnologies. A.W.R. declares institutional support from the Steven and Alexandra Cohen Foundation, and Global Lyme Alliance. M.J.S. declares institutional support from the Steven and Alexandra Cohen Foundation, and NIH grant P30 AR070254. E.J.H. declares compensation from the Lyme Disease Biobank and institutional support from Bay Area Lyme Foundation, the Steven and Alexandra Cohen Foundation, and Adaptive Biotechnologies. R.O.E. and T.M. declare employment with Adaptive Biotechnologies during the time of this research. L.B. and H.S.R. declare employment, equity ownership and leadership with Adaptive Biotechnologies. J.N.A. declares consulting with Tarsus Pharmaceuticals and Pfizer; participation in an advisory board with Adaptive Biotechnologies; expert testimony; membership on the Bay Area Lyme Foundation Scientific Advisory Board; Past Chair, 2018, HHS Tick-borne Disease Working Group, Office of HIV/AIDS and Infectious Disease Policy, Office of the Assistant Secretary of Health, Department of Health and Human Services; and institutional support from the Steven and Alexandra Cohen Foundation and Global Lyme Alliance.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

REFERENCES

1. Skar GL, Simonsen KA. 2021. Lyme Disease. 2021. In StatPearls, StatPearls Publishing, Treasure Island, FL.
2. Schwartz AM, Kugeler KJ, Nelson CA, Marx GE, Hinckley AF. 2021. Use of commercial claims data for evaluating trends in Lyme disease diagnoses, United States, 2010–2018. Emerg Infect Dis 27:499–507.
3. Kugeler KJ, Schwartz AM, Delorey MJ, Mead PS, Hinckley AF. 2021. Estimating the frequency of Lyme disease diagnoses, United States, 2010–2018. Emerg Infect Dis 27:616–619.
4. Wormser GP, Dattwyler RJ, Shapiro ED, Halperin JJ, Steere AC, Klempner MS, Krause PJ, Bakken JS, Strle F, Stanek G, Bockenstedt L, Fish D, Dumler JS, Nadelman RB. 2006. The clinical assessments treatment, and prevention of Lyme disease, human granulocytic anaplasmosis, and babesiosis: clinical practice guidelines by the Infectious Diseases Society of America. Clin Infect Dis 43:1089-134.
5. Center for Disease Control and Prevention. 2021. Lyme and other tickborne diseases increasing. https://www.cdc.gov/media/dpk/diseases-and-conditions/lyme-disease/index.html. Accessed 1 July 2021.
6. Alkishe A, Raghavan RK, Peterson AT. 2021. Likely geographic distributional shifts among medically important tick species and tick-associated diseases under climate change in North America: A review. Insects 12:225.
7. Steere AC. 2001. Lyme disease. N Engl J Med 345:115–125.
8. Makhani N, Morris SK, Page A V., Brophy J, Lindsay LR, Banwell BL, Richardson SE. 2011. A

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

twist on lyme: The challenge of diagnosing european lyme neuroborreliosis. J Clin Microbiol 49:455–457.

9. Hengge UR, Tannapfel A, Tyring SK, Erbel R, Arendt G, Ruzicka T. 2003. Lyme borreliosis. Lancet Infect Dis 3:489-500.

10. Stanek G, Wormser GP, Gray J, Strle F. 2012. Lyme borreliosis. Lancet 379:461-473.

11. Torbahn G, Hofmann H, Rücker G, Bischoff K, Freitag MH, Dersch R, Fingerle V, Motschall E, Meerpohl JJ, Schmucker C. 2018. Efficacy and safety of antibiotic therapy in early cutaneous Lyme borreliosis: a network meta-analysis. JAMA Dermatol 154:1292–1303.

12. Center for Disease Control and Prevention. 2011. Two-tiered testing decision tree. https://www.cdc.gov/lyme/healthcare/clinician_twotier.html. Accessed 11 May 2021.

13. Hatchette T, Davis I, Johnston B. 2014. Clinical aspects of Lyme disease: Lyme disease: clinical diagnosis and treatment. Canada Commun Dis Rep 40:194-208.

14. Murray TS, Shapiro ED. 2010. Lyme disease. Clin Lab Med 30:311-328.

15. Smith RP, Schoen RT, Rahn DW, Sikand VK, Nowakowski J, Parenti DL, Holman MS, Persing DH, Steere AC. 2002. Clinical characteristics and treatment outcome of early lyme disease in patients with microbiologically confirmed erythema migrans. Ann Intern Med 136:421–428.

16. Lantos PM, Rumbaugh J, Bockenstedt LK, Falck-Ytter YT, Aguero-Rosenfeld ME, Auwaerter PG, Baldwin K, Bannuru RR, Belani KK, Bowie WR, Branda JA, Clifford DB, DiMario FJ, Halperin JJ, Krause PJ, Lavergne V, Liang MH, Cody Meissner H, Nigrovic LE, Nocton J (Jay) J, Osani MC, Pruitt AA, Rips J, Rosenfeld LE, Savoy ML, Sood SK, Steere AC, Strle F, Sundel R, Tsao J, Vaysbrot EE, Wormser GP, Zemel LS. 2021. Clinical practice guidelines by the

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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. Arthritis Care Res 73:1–9.

17. Charbonneau A, Charette L-P, Rouleau G, Savary M, Wilson A, Heer E, Bériault K, de Pokomandy A. 2018. Clinical presentation of Lyme disease in the higher-risk region of Quebec: a retrospective descriptive study. CMAJ Open 6:E139–E145.
18. Santino I, Longobardi V. 2011. Clinical and serological features of patients with suspected Lyme borreliosis. Int J Immunopathol Pharmacol 24:797–801.
19. Center for Disease Control and Prevention. 1995. Recommendations for test performance and interpretation from the Second National Conference on Serologic Diagnosis of Lyme Disease. MMWR Morb Mortal Wkly Rep 44:590–591.
20. Sigal LH. 1995. Diagnosis of Lyme disease. JAMA 274:1427-1428.
21. Marques AR. 2018. Revisiting the lyme disease serodiagnostic algorithm: the momentum gathers. J Clin Microbiol 56:e00749-18.
22. Seriburi V, Ndukwe N, Chang Z, Cox ME, Wormser GP. 2012. High frequency of false positive IgM immunoblots for Borrelia burgdorferi in Clinical Practice. Clin Microbiol Infect 18:1236–1240.
23. Nadelman RB, Nowakowski J, Forseter G, Goldberg NS, Bittker S, Cooper D, Aguero-Rosenfeld M, Wormser GP. 1996. The clinical spectrum of early Lyme borreliosis in patients with culture-confirmed erythema migrans. Am J Med 100:502–508.
24. Conant JL, Powers J, Sharp G, Mead PS, Nelson CA. 2018. Lyme disease testing in a high-incidence state. Am J Clin Pathol 149:234–240.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

25. Fix AD, Strickland GT, Grant J. 1998. Tick bites and lyme disease in an endemic: Setting problematic use of serologic testing and prophylactic antibiotic therapy. JAMA 279:206–210.

26. Nowakowski J, Schwartz I, Liveris D, Wang G, Aguero-Rosenfeld MEA, Girao G, McKenna D, Nadelman RB, Cavaliere LF, Wormser GP. 2001. Laboratory diagnostic techniques for patients with early Lyme disease associated with erythema migrans: a comparison of different techniques. Clin Infect Dis 33:2023–2027.

27. Schriefer ME. 2015. Lyme disease diagnosis: serology. Clin Lab Med 35:797-814.

28. Steere AC, McHugh G, Damle N, Sikand VK. 2008. Prospective study of serologic tests for Lyme disease. Clin Infect Dis 47:188–195.

29. Mead P, Petersen J, Hinckley A. 2019. Updated CDC recommendation for serologic diagnosis of Lyme disease. MMWR Morb Mortal Wkly Rep 68:703.

30. Zweitzig D, Kopnitsky M, ZEUS Scientific. Validation of a modified two-tiered testing (MTTT) algorithm for the improved diagnosis of Lyme disease. ZEUS Scientific, Branchburg, NJ.

31. Branda JA, Strle K, Nigrovic LE, Lantos PM, Lepore TJ, Damle NS, Ferraro MJ, Steere AC. 2017. Evaluation of modified 2-tiered serodiagnostic testing algorithms for early Lyme disease. Clin Infect Dis 64:1074–1080.

32. Bunikis J, Barbour AG. 2002. Laboratory testing for suspected Lyme disease. Med Clin North Am 86:311–340.

33. Rebman AW, Crowder LA, Kirkpatrick A, Aucott JN. 2015. Characteristics of seroconversion and implications for diagnosis of post-treatment Lyme disease syndrome: acute and

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

convalescent serology among a prospective cohort of early Lyme disease patients. Clin Rheumatol 34:585–589.

34. Hinckley AF, Connally NP, Meek JI, Johnson BJ, Kemperman MM, Feldman KA, White JL, Mead PS. 2014. Lyme disease testing by large commercial laboratories in the United States. Clin Infect Dis 59:676–681.

35. Dressier F, Yoshinari NH, Steere AC. 1991. The T-Cell proliferative assay in the diagnosis of Lyme disease. Ann Intern Med 115:533–539.

36. Vaz A, Glickstein L, Field JA, McHugh G, Sikand VK, Damle N, Steere AC. 2001. Cellular and humoral immune responses to Borrelia burgdorferi antigens in patients with culture-positive early Lyme disease. Infect Immun 69:7437-7444.

37. Soloski MJ, Crowder LA, Lahey LJ, Wagner CA, Robinson WH, Aucott JN. 2014. Serum inflammatory mediators as markers of human Lyme disease activity. PLoS One 9:e93243.

38. Hemmer B, Gran B, Zhao Y, Marques A, Pascal J, Tzou A, Kondo T, Cortese I, Bielekova B, Straus SE, Mcfarland HF, Houghten R, Simon R, Pinilla C, Martin R. 1999. Identification of candidate T-cell epitopes and molecular mimics in chronic Lyme disease. Nat Med 5:1375–1382.

39. Steere AC. 2019. Treatment of Lyme arthritis. J Rheumatol 46:871-873.

40. Gross DM, Steere AC, Huber BT. 1998. T helper 1 response is dominant and localized to the synovial fluid in patients with Lyme arthritis. J Immunol 160:1022–1028.

41. Kalish R, McHugh G, Granquist J, Shea B, Ruthazer R, Steere A. 2001. Persistence of immunoglobulin M or immunoglobulin G antibody responses to Borrelia burgdorferi 10-20 years after active Lyme disease. Clin Infect Dis 33:780–785.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

42. Venturi V, Price DA, Douek DC, Davenport MP. 2008. The molecular basis for public T-cell responses? Nat Rev Immunol 8:231-238.

43. Emerson RO, DeWitt WS, Vignali M, Gravley J, Hu JK, Osborne EJ, Desmarais C, Klinger M, Carlson CS, Hansen JA, Rieder M, Robins HS. 2017. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. Nat Genet 49:659–665.

44. Snyder M T, Gittelman M R, Klinger M, May H D, Osborne J E, Taniguchi R, Jabran Zahid H, Kaplan M I, Dines N J, Noakes T M, Pandya R, Chen X, Elasady S, Svejnoha E, Ebert P, Pesesky W M, De Almeida P, O'Donnell1 H, DeGottardi Q, Keitany G, Lu J, Vong A, Elyanow R, Fields P, Greissl J, Baldo1 L, Semprini S, Cerchione C, Nicolini F, Mazza M, Delmonte M. Ottavia, Dobbs K, Laguna-Goya R, Carreño-Tarragona G, Barrio S, Imberti L, Sottini A, Quiros-Roldan E, Rossi C, Biondi A, Bettini R L, D'Angio M, Bonfanti P, Tompkins F. Miranda, Alba C, Dalgard C, Sambri V, Martinelli G, Goldman D J, Heath R J, Su C H, Notarangelo D L, Paz-Artal E, Martinez-Lopez J, Carlson M. Jonathan, Robins S H. 2020. Magnitude and dynamics of the T-cell response to SARS-CoV-2 infection at both individual and population levels. medRxiv https://doi.org/10.1101/2020.07.31.20165647.

45. Gittelman RM, Lavezzo E, Snyder TM, Jabran Zahid H, Elyanow R, Dalai S, Kirsch I, Baldo L, Manuto L, Franchin E, Vecchio C Del, Simeoni F, Bordini J, Lorè NI, Lazarevic D, Cirillo DM, Ghia P, Toppo S, Carlson JM, Robins HS, Tonon G, Crisanti A. 2020. Diagnosis and tracking of past SARS-CoV-2 infection in a large study of Vo', Italy through T-cell receptor sequencing. medRxiv https://doi.org/10.1101/2020.11.09.20228023.

46. Elyanow R, Snyder TM, Dalai SC, Gittelman RM, Boonyaratanakornkit J, Wald A, Selke S,

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Wener MH, Morishima C, Greninger AL, Holbrook MR, Kaplan IM, Zahid HJ, Carlson JM, Baldo L, Manley T, Robins HS, Koelle DM. 2021. T-cell receptor sequencing identifies prior SARS-CoV-2 infection and correlates with neutralizing antibody titers and disease severity. medRxiv https://doi.org/10.1101/2021.03.19.21251426.

47. Dalai SC, Dines JN, Snyder TM, Gittelman RM, Eerkes T, Vaney P, Howard S, Akers K, Skewis L, Monteforte A, Witte P, Wolf C, Nesse H, Herndon M, Qadeer J, Duffy S, Svejnoha E, Taromino C, Kaplan IM, Alsobrook J, Manley T, Baldo L. 2021. Clinical validation of a novel T-cell receptor sequencing assay for identification of recent or prior SARS-CoV-2 infection. medRxiv https://doi.org/10.1101/2021.01.06.21249345.

48. Mosel M, Rebman A, Carolan H, Montenegro T, Lovari R, Schutzer S, Ecker D, Yang T, Ramadoss N, Robinson W, Soloski M, Eshoo M, Aucott J. 2020. Molecular microbiological and immune characterization of a cohort of patients diagnosed with early Lyme disease. J Clin Microbiol 59:e00615-20.

49. Aucott JN, Rebman AW, Crowder LA, Kortte KB. 2013. Post-treatment Lyme disease syndrome symptomatology and the impact on life functioning: is there something here? Qual Life Res 22:75–84.

50. Zajkowska JM. 2014. Antibody-based techniques for detection of Lyme disease: a challenging issue. Antib Technol J 4:33–44.

51. Nolan S, Vignali M, Klinger M, Dines J, Kaplan I, Svejnoha E, Craft T, Boland K, Pesesky M, Gittelman R, Snyder T, Gooley C, Semprini S, Cerchione C, Mazza M, Delmonte O, Dobbs K, Carreño-Tarragona G, Barrio S, Sambri V, Martinelli G, Goldman J, Heath J, Notarangelo L, Carlson J, Martinez-Lopez J, Robins H. 2020. A large-scale database of T-cell receptor beta

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(TCRβ) sequences and binding associations from natural and synthetic exposure to SARS-CoV-2. Res Sq https://doi.org/10.21203/rs.3.rs-51964/v1

52. McKisic MD, Barthold SW. 2000. T-cell-independent responses to Borrelia burgdorferi are critical for protective immunity and resolution of Lyme disease. Infect Immun 68:5190-5197.

53. Hastey CJ, Elsner RA, Barthold SW, Baumgarth N. 2012. Delays and diversions mark the development of B cell responses to Borrelia burgdorferi infection. J Immunol 188:5612-5622.

54. Melia MT, Auwaerter PG. 2016. Time for a different approach to Lyme disease and long-term symptoms. NEJM 374:1277–1278.

55. Aucott JN, Soloski MJ, Rebman AW, Crowder LA, Lahey LJ, Wagner CA, Robinson WH, Bechtold KT. 2016. CCL19 as a chemokine risk factor for posttreatment Lyme disease syndrome: a prospective clinical cohort study. Clin Vaccine Immunol 23:757-766.

56. Strle K, Stupica D, Drouin EE, Steere AC, Strle F. 2014. Elevated levels of IL-23 in a subset of patients with post–Lyme disease symptoms following erythema migrans. Clin Infect Dis 58:372-380.

57. Horn E, Dempsey G, Schotthoefer A, Prisco U, McArdle M, Gervasi S, Golightly M, De Luca C, Evans M, Pritt B, Theel E, Iyer R, Liveris D, Wang G, Goldstein D, Schwartz I. 2020. The Lyme Disease Biobank: Characterization of 550 patient and control samples from the East Coast and Upper Midwest of the United States. J Clin Microbiol 58:e00032-20.

58. Rebman AW, Yang T, Mihm EA, Novak CB, Yoon I, Powell D, Geller SA, Aucott JN. 7 Mar 2021. The presenting characteristics of erythema migrans vary by age, sex, duration, and

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

body location. Infection 10.1007/s15010-021-01590-0.

59. Aucott J, Crowder L, Kortte K. 2013. Development of a foundation for a case definition of post-treatment Lyme disease syndrome. Int J Infect Dis 17:e443-449.

60. Nelson WC, Pyo CW, Vogan D, Wang R, Pyon YS, Hennessey C, Smith A, Pereira S, Ishitani A, Geraghty DE. 2015. An integrated genotyping approach for HLA and other complex genetic systems. Hum Immunol 76:928–938.

61. Smith AG, Pereira S, Jaramillo A, Stoll ST, Khan FM, Berka N, Mostafa AA, Pando MJ, Usenko CY, Bettinotti MP, Pyo C-W, Nelson WC, Willis A, Askar M, Geraghty DE. 2019. Comparison of sequence-specific oligonucleotide probe vs next generation sequencing for HLA-A, B, C, DRB1, DRB3/B4/B5, DQA1, DQB1, DPA1, and DPB1 typing: toward single-pass high-resolution HLA typing in support of solid organ and hematopoietic cell transplant programs. HLA 94:296–306.

62. Robins HS, Campregher P V., Srivastava SK, Wacher A, Turtle CJ, Kahsai O, Riddell SR, Warren EH, Carlson CS. 2009. Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. Blood 114:4099–4107.

63. Robins H, Desmarais C, Matthis J, Livingston R, Andriesen J, Reijonen H, Carlson C, Nepom G, Yee C, Cerosaletti K. 2012. Ultra-sensitive detection of rare T cell clones. J Immunol Methods 375:14–19.

64. Carlson CS, Emerson RO, Sherwood AM, Desmarais C, Chung MW, Parsons JM, Steen MS, LaMadrid-Herrmannsfeldt MA, Williamson DW, Livingston RJ, Wu D, Wood BL, Rieder MJ, Robins H. 2013. Using synthetic templates to design an unbiased multiplex PCR assay. Nat Commun 4:2680.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

65. Pavlović M, Scheffer L, Motwani K, Kanduri C, Kompova R, Vazov N, Waagan K, Bernal FLM, Costa AA, Corrie B, Akbar R, Hajj GS Al, Balaban G, Brusko TM, Chernigovskaya M, Christley S, Cowell LG, Frank R, Grytten I, Gundersen S, Haff IH, Hochreiter S, Hovig E, Hsieh P-H, Klambauer G, Kuijjer ML, Lund-Andersen C, Martini A, Minotto T, Pensar J, Rand K, Riccardi E, Robert PA, Rocha A, Slabodkin A, Snapkov I, Sollid LM, Titov D, Weber CR, Widrich M, Yaari G, Greiff V, Sandve GK. 2021. immuneML: an ecosystem for machine learning analysis of adaptive immune receptor repertoires. bioRxiv https://doi.org/10.1101/2021.03.08.433891

66. Cannon M, Schmid D, Hyde T. 2010. Review of cytomegalovirus seroprevalence and demographic characteristics associated with infection. Rev Med Virol 20:202–213.

67. Klinger M, Pepin F, Wilkins J, Asbury T, Wittkop T, Zheng J, Moorhead M, Faham M. 2015. Multiplex identification of antigen-specific T cell receptors using a combination of immune assays and immune receptor sequencing. PLoS One 10:e0141561.

68. Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res 14:1188–1190.

69. Schneider T, Stephens R. 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res 18:6097–6100.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

FIGURE LEGENDS

Figure 1: Lyme disease-associated TCRs distinguish cases from controls in training cohorts. (A) Distribution of the number of TCR rearrangements encoding enhanced sequences as a function of the (log-transformed) total number of unique TCR rearrangements identified in a repertoire. The distribution of enhanced sequences in control samples approximately follows a logistic growth curve (solid black line; dashed red lines indicate +2, +3, and +4 standard deviations from fit), which was used to define a scoring function. Case (orange; *n* = 72) and control (blue; *n* = 2,981) samples from the training set (see Table S1) are shown. (B) Distribution of the resulting model score is largely invariant to the number of unique rearrangements in an independent set of endemic controls (*n* = 2,627). Red line indicates 99th percentile distribution in this cohort (score = 4.2675), which was used to define the positive call threshold. (C) Model score distribution in a holdout set of repertoires from the LDB cohort (*n* = 15 cases and 48 controls), collected in the 2019 tick season and immunosequenced after model training.

Figure 2: Validation of the TCR classifier in the JHU cohort and other holdout endemic controls. (A) Model score distribution in early Lyme disease samples from JHU (blue, *n* = 211), in addition to holdout endemic controls. JHU (orange, *n* = 45) and LDB (green, *n* = 116) endemic controls were deemed to not have Lyme disease based on clinical assessment and negative STTT testing. Other endemic controls (red, *n* = 2516) were drawn from our database of repertoires sampled from individuals in endemic regions in the US and Europe who were presumed negative for Lyme disease. (B) Sensitivity of STTT and the TCR classifier for individuals

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

in the JHU cohort, stratified by symptom duration (days) at time of enrollment. Participants were stratified based on self-reported symptom duration. Error bars represent mean ± 95% CI by bootstrap sampling. (C) Model score distribution for JHU early disease samples stratified by STTT serostatus at enrollment and posttreatment follow-up. Positive (blue, *n* = 64): STTT-positive at enrollment; Converter (orange, *n* = 38): STTT-negative at enrollment and STTT-positive at posttreatment follow-up; Negative (green, *n* = 109): STTT-negative at both visits. Boxes indicate median ± interquartile ranges (IQR), and whiskers denote 1.5 times the IQR above the high quartile and below the low quartile. (D) Receiver operating characteristic (ROC) curves using all endemic control samples from (A) as negatives. Areas under the ROC curves are 0.98, 0.89, and 0.71 for the Positive, Converter, and Negative curves, respectively.

Figure 3: Longitudinal dynamics of TCR scoring by initial serostatus for the JHU cohort. TCR repertoires were available from timepoints before (0w) and after (3w) treatment and 6 months posttreatment (6mo) for 161 of the JHU participants. Positive (blue, *n* = 53): STTT-positive at enrollment; Converter (orange, *n* = 32): STTT-negative at enrollment and STTT-positive at posttreatment follow-up; Negative (green, *n* = 76): STTT-negative at both visits. Boxes indicate median ± IQR, and whiskers denote 1.5 times the IQR above the high quartile and below the low quartile.

Figure 4: Clinical correlates of TCR model score. TCR scores were stratified by a) liver function test results (elevated [*n* = 72] versus normal [*n* = 139]), b) lymphocyte counts (normal [*n* = 150]

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

versus low [*n* = 61]), or c) presentation of rash (multiple/disseminated [*n* = 68] versus single [*n* = 143]); and d) plotted as a function of the number of Lyme-related symptoms (Spearman R^2 = 0.17. *P*-values, Mann-Whitney U test). Boxes indicate median ± IQR, and whiskers denote 1.5 times the IQR above the high quartile and below the low quartile.

Figure 5: Percentage of samples with enhanced sequences (ES) assigned by MIRA to the indicated *B. burgdorferi* antigens from early Lyme cases from the JHU cohort and endemic controls. FlaB, flagellin protein B; DbpA, decorin-binding protein A.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

TABLES

Table 1: TCR classifier sensitivity stratified by serostatus. Ranges indicate 95% bootstrap CI.

Table 2: TCR clusters as defined by connected components within a 1–amino-acid change in CDR3 and the same V-gene family.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

SUPPLEMENTARY MATERIALS

Supplementary Figure S1: Immunosequencing input DNA distributions by cohort. Boxes indicate median ± interquartile ranges (IQR), and whiskers denote 1.5 times the IQR above the high quartile and below the low quartile.

Supplementary Table S1: Cohorts used for training, setting the classification threshold, and validation.

Supplementary Table S2: Multiple logistic regression of TCR model score on clinical features.

Supplementary Table S3: Counts of enhanced sequences mapped to each protein by MIRA.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

METHODS

Ethics

All procedures involving human participants were conducted in accordance with the ethical standards of the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. For the Lyme Disease Biobank (LDB) cohort, Institutional review board (IRB) approval was obtained for each site through the LDB sponsor protocol (Advarra IRB) or institution-specific IRB. For the Boca Biolistics cohort, IRB approval was obtained through Advarra IRB. Human subjects protocols for the JHU cohort were approved by the IRBs of Johns Hopkins University and Stanford University. All participants provided written informed consent prior to enrollment.

Study cohorts

Overview

The repertoires used in this study were sampled from 8,590 donors enrolled in multiple studies, described in detail below. As the aim of this study was to develop and validate a model for identifying early Lyme disease with high sensitivity and specificity, allocation of samples to sets used for training, setting the classification threshold, and validation was prespecified (see Table S1 for detailed cohort information).

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Lyme Disease Biobank (LDB) cohort

LDB is a program of the Bay Area Lyme Foundation. The LDB cohort enrolled individuals from the East coast and upper Midwest regions of the US who presented with signs or symptoms consistent with early Lyme disease. Patients presenting with erythema migrans (EM) rash >5 cm or an erythematous, annular, expanding skin lesion ≤5 cm or presenting with signs or symptoms (headache, fatigue, fever, chills, or joint or muscular pain) without an EM/annular lesion, but with a suspected tick exposure or tick bite, and with no history of chronic fatigue syndrome, rheumatologic disease, or multiple sclerosis, were included. Individuals with tick-bite reactions (e.g., a nonannular erythematous macule at the site of the tick bite) without EM or expanding annular lesion were excluded, as were those who had initiated antibiotics >48 hours prior to enrollment. Healthy individuals living in the same regions with no history of Lyme disease or tick-borne infection were enrolled as endemic controls. Real-time PCR for *B. burgdorferi* and 2-tiered standard testing (STTT) was conducted on all samples including controls. Laboratory-confirmed early Lyme disease samples were defined as being STTT-positive, PCR-positive (sample or culture), or having 2 positive enzyme-linked immunosorbent assays and an EM ≥5 cm. Laboratory-confirmed endemic controls were defined as being STTT-negative. Additional details and baseline clinical characteristics of this sample have been previously published (57). Only laboratory-confirmed samples that were STTT-positive were used in model training. Only laboratory-confirmed–positive (by STTT, PCR, or culture) cases and laboratory-confirmed–negative (by STTT) endemic controls were included in analyses of model performance or to establish the final call threshold.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Boca Biolistics cohort

Specimens were collected from antibiotic treatment-naïve patients recruited at clinical sites throughout New York and New Jersey who presented with acute symptomology of a tick-borne illness. Participants had blood drawn on 3 occasions: ≤30 days post tick bite while antibiotic treatment naïve, 6 to 8 weeks post tick bite, and 16 to 24 weeks post tick bite. Whole blood samples were aliquoted, frozen, and stored at −80 °C after collection. At each visit, information was captured regarding the individual's symptoms and date of symptom onset, treatment status, treatment regimen, and standard-of-care lab results. Once the specimens were received at Boca Biolistics, they were characterized for all relevant tick-borne pathogens including *B. burgdorferi*, *B. microti*, *E. chaffeensis*, and *A. phagocytophilum*. Testing was performed both in house and at ARUP Laboratories on matched serum collected from donors. DiaSorin and Immunonetics Lyme antibody testing was performed at Boca Biolistics Reference Laboratory, and IgM- and IgG-specific antibody screening for *B. burgdorferi*, *B. microti*, *E. chaffeensis*, and *A. phagocytophilum* was performed at ARUP Laboratories. For the present study, immunosequencing was performed on the first available sample from 18 donors who were seropositive for *B. burgdorferi* by enzyme-linked immunosorbent assay (ELISA) and immunoblot (either IgG or IgM), all of whom were classified as STTT-positive by 2-tiered testing criteria.

Johns Hopkins University (JHU) cohort

Study of Lyme disease Immunology and Clinical Events (SLICE) was a longitudinal, prospective cohort study that enrolled adult patients (≥18 years of age) with early Lyme disease who were

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

self-referred or recruited from primary or urgent care settings from 2008 to 2020. Eligible participants were primarily enrolled at study sites in Maryland, with a small number enrolled at a satellite site in southeastern Pennsylvania. At enrollment, participants were required to have a visible EM ≥5 cm in diameter diagnosed by a health care provider and either multiple skin lesions or at least one new-onset concurrent symptom. All patients had received ≤72 hours of appropriate antibiotic treatment for early Lyme disease at enrollment. Additional details and baseline clinical characteristics of this sample have been previously published (58). Participants without a clinical or serologic history of Lyme disease were recruited from similar primary care settings or through the community using flyers and online advertising to serve as endemic controls. This cohort was required to be STTT-negative at the time of enrollment and at all subsequent visits, as well as be free of any history of prior clinical Lyme disease. All participants in both groups were excluded for a range of self-reported prior medical conditions paralleling those listed in the proposed case definition for posttreatment Lyme disease syndrome (PTLDS) (59), specifically chronic fatigue syndrome, fibromyalgia, unexplained chronic pain, sleep apnea or narcolepsy, autoimmune disease, chronic neurologic disease, liver disease, hepatitis, HIV, cancer or malignancy in the past 2 years, major psychiatric illness, or drug or alcohol abuse.

All patients were treated with 3 weeks of oral doxycycline in accordance with IDSA Guidelines (16). Lyme patients were seen regularly over the course of 2 years for a total of 5 study visits (before and immediately after treatment, and 1 month, 6 months, and 2 years posttreatment). Samples collected before and immediately after treatment and 6 months posttreatment were used for the present study. Control samples from healthy individuals were collected at an initial visit and 6 months and 1 year later; samples from the initial visit were used in the present

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

study. Disseminated EM rash was defined as having more than one visible rash site, while local rash was defined as a single EM rash site.

High-resolution HLA class I and class II typing for the JHU cohort (cases only) was performed by Scisco Genetics, Inc., (Seattle, WA, USA) using the ScisGo HLA v6 typing kit, as previously described (60, 61).

Database controls

A total of 7,964 repertoires that were sampled as part of previous studies were selected from our database. Inclusion was determined at the cohort level and based on the size of the cohort, geographic region (US and Lyme-endemic regions of Europe), and sequencing date (2019 or later, to ensure consistent lab sequencing protocols). These repertoires were from individuals defined as being either from endemic regions (cohorts from Germany, Italy, or upper Midwest or Northeast regions of US) or non-endemic regions (other regions of the US). All individuals in these cohorts were presumed to be Lyme negative but were not tested for Lyme disease.

Assignment of cohorts

All training cases were drawn from the LDB and Boca Biolistics cohorts. To enrich for cases with a likely immune response in order to maximize our ability to detect Lyme disease-associated enhanced sequences, the training set was limited to 72 STTT-positive cases (54 from LDB, 18 from Boca Biolistics). Training controls included 2,981 repertoires from individuals from non-

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

endemic regions of the US and Europe that were previously collected as a part of other studies and presumed to have never been exposed to *B. burgdorferi* infection.

The positive-call threshold was set based on 2,507 presumed Lyme-negative samples collected from endemic regions that were available in our database, along with 120 confirmed STTT-negative endemic controls randomly selected from the LDB cohort. Additional LDB case (n = 15) and control (n = 48) samples collected during the 2019 tick season were sequenced after model training and used as an initial check of model specificity and generalizability.

The primary endpoint of the study was evaluation of sensitivity in the JHU cohort, which was selected based on the conservative enrollment criteria for that cohort. Repertoires sampled from 211 participants at time of enrollment passed quality control (QC) thresholds established after model training described below. A subset of patients in the JHU cohort (n = 161) had sequenced repertoires that passed QC from samples collected before and after treatment and 6 months posttreatment.

Specificity of the final model was estimated based on 3 endemic control cohorts: 1) all endemic controls from JHU (n = 45); 2) 50% of endemic controls from tick seasons prior to 2019 in the LDB cohort (selected by random sampling; n = 115 passed QC); and 3) 50% of presumed endemic controls from our database (selected by random sampling; n = 2,471 passed QC).

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Immunosequencing of T-cell receptor (TCR) repertoires

Immunosequencing of complementarity-determining region 3 (CDR3) of human TCRβ chains was performed using the immunoSEQ® Assay (Adaptive Biotechnologies, Seattle, WA). Extracted genomic DNA was amplified in a bias-controlled multiplex PCR, followed by high-throughput sequencing. Sequences were collapsed and filtered in order to identify and quantitate the absolute abundance of each unique TCRβ CDR3 region for further analysis as previously described (62–64). Sequencing reactions contained a median of 7,884.0 ng of input DNA (range, 239.9 to 55,186.4 ng) and yielded a median of 314,948 T-cell templates per sample (range, 15 to 1,837,496) (Fig. S1). The T-cell fraction (percentage of T cells among the estimated number of nucleated cells input) ranged from 0.3% to 90%, with a median of 26.0%.

Specification of the TCR-based Lyme classification model

Identification of Lyme disease-associated enhanced sequences

Public TCRβ amino acid sequences associated with early Lyme disease were identified as described previously (43). Briefly, one-tailed Fisher's exact tests (FETs) were performed on all unique TCR sequences to compare frequencies in early Lyme samples with those in presumed-negative controls. Unique sequences were defined based on the V gene, J gene, and CDR3 amino acid sequence. The *P*-value threshold for including a TCR in the enhanced sequence list was treated as a hyperparameter and was selected to maximize model performance as

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

described below. The resulting set of FET-defined enhanced sequences for cohort $\mathcal{C}$ are denoted $S_{fet}^{\mathcal{C}}$.

Other features of a TCR sequence that are indicative of clonal expansion can help identify additional disease-associated sequences that are missed when considering only the presence or absence of a sequence during identification of FET-defined enhanced sequences. These features include 1) the convergent recombination (CR) count (defined as the number of unique nucleotide TCR clones encoding TCRβ amino acid sequences observed in a repertoire), 2) the productive frequency of a sequence in a repertoire; and 3) the tendency of a TCR to cluster with other enhanced sequences (defined here as the number of sequences in S_{fet} that are similar to the TCR, defined as sharing the same V gene, having identical CDR3 length, and differing by one amino acid). To identify additional disease associated TCRs, it is ideal to build a model that can weight these additional features appropriately.

A recent study in CMV has demonstrated that many TCRs that are not identified as significant enhanced sequences in small training datasets based solely on FET may be selected in larger training sets (65). This observation motivates a simple classification problem: prediction of whether a TCR will be identified as an enhanced sequence by FET when the dataset grows to a specified size. To this end, a logistic regression model was fitted, where the training data was the set of TCRs and corresponding features observed in a previously reported cohort $\mathcal{C}_{CMV}$ labeled for CMV serostatus (43). The dependent binary variable was defined as 1 if the TCR was observed as in in $S_{fet}^{\mathcal{C}^*}$ for some large CMV-labeled cohort C_{CMV}^*, and 0 otherwise. For each TCR, the following features were defined: 1) average and maximum CR for cases and controls; 2)

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

average and maximum productive frequency for cases and controls; and 3) the number of sequences in S_{fet} that are similar to the TCR, as defined above.

In practice, a larger CMV-labeled cohort was unavailable. However, as >50% of North American and European populations are seropositive for CMV (66), we applied a simple pseudolabeling procedure to construct C^*_{CMV}. Briefly, a logistic regression classifier (as defined in [44]) based on $S_{fet}^{\mathcal{C}_{CMV}}$ was trained on the labeled CMV cohort $\mathcal{C}_{CMV}$, then applied to all samples in the Lyme training cohort. The inferred CMV status was then treated as observed and combined with $\mathcal{C}_{CMV}$, resulting in $\mathcal{C}^*_{CMV}$, which was used to define $S_{fet}^{\mathcal{C}^*_{CMV}}$. The resulting logistic regression classifier was able to accurately predict which TCRs observed in $\mathcal{C}_{CMV}$ but not in $S_{fet}^{\mathcal{C}_{CMV}}$ would end up in $S_{fet}^{\mathcal{C}^*_{CMV}}$ (area under the reviewer operating characteristic [ROC] curve = 0.84 in cross validation; data not shown). The features receiving the greatest weight in this model were the convergent recombination counts in cases (likely indicating substantial clonal expansion) and the number of similar sequences in $S_{fet}^{\mathcal{C}_{CMV}}$ (likely indicating that the TCR responds to the same antigen as another enhanced sequence).

The model fitted to CMV was used to infer enhanced sequences for Lyme disease. Combining these inferred enhanced sequences with enhanced sequences identified by Fisher's exact test resulted in the final set of Lyme disease-associated enhanced sequences, $S^{\mathcal{C}_{Lyme}}$. Hyperparameters in this model were chosen using cross-validation in the context of the disease classification model described below.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Inferring early Lyme disease status based on enhanced sequence counts in TCR repertoires

Given a set of enhanced sequences S, the pair (y_i, x_i) can then be defined for each repertoire i, where x_i is the total number of unique productive DNA TCR rearrangements in the sampled repertoire, and $y_i < x_i$ is the number of those rearrangements that encode any of the enhanced sequences in S. If y_i is treated as sampled from a random variable Y, the expected value of Y given x can be considered. By the way enhanced sequences are defined, the distribution of $Y|x$ is expected to vary substantially between cases and controls. While this could be treated as a classification problem to maximize the separation between cases and controls (as in [43, 44]), $Y|x$ was instead explicitly modelled among control samples, with classification based on standard units of deviation above and below expectation. This approach provides superior control of specificity across populations given the extremely unbalanced nature of our case/control data set. To model this distribution, Y was assumed to follow a binomial distribution, with mean $f(x) = y_{\max}\, p(x)$ and variance $\sigma^2(x) = y_{\max}\, p(x)(1 - p(x))$, where $y_{\max}$ is the maximum number of enhanced sequences observed in any training repertoire, and

$$p(x) = \frac{1}{1 + \exp(-(w \log_{10} x + b))}$$

for model parameters w and b. For a given (y_i, x_i), the number of standard deviations y_i is from the expected mean given x_i is then used as the model score:

$$ModelScore(y_i, x_i) = \frac{y_i - f(x_i)}{\sigma(x_i)}.$$

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The model parameters w and b were chosen by minimizing the sum of squared residuals over the set of training control samples.

The observed data are moderately overdispersed with respect to the estimated variance (Fig 1A). As such, the final call threshold t was chosen to fix the prespecified false-positive rate of 1% on a set of 2,627 presumed Lyme-negative control samples, as described above.

TCR repertoire quality control (QC) criteria

The two key parameters of the classifier are the number of unique productive rearrangements, x, and the number of unique productive rearrangements encoding an enhanced sequence, y. For a given blood sample, the value of x is determined by the quantity of DNA, the fraction of cells that are T-cells, and the diversity of T-cells. In rare cases, x is too small to yield meaningful information, or significantly larger than observed in our training data, making extrapolation of $Y|x$ problematic. Therefore, acceptance criteria were predefined for the number of unique productive rearrangements based on the observed distribution of x in the training data.

The information contained in enhanced sequences is asymmetric: for small x, large y is considered to be evidence of Lyme disease, while small y may simply reflect a lack of sequenced T cells. Thus, QC criteria were treated asymmetrically. Specifically, $x_{\max}$ and $x_{\min}$ were defined as the upper and lower QC thresholds, which were prespecified to be equal to the 1st and 99th percentiles, respectively, of x observed in the training data. A sample i then failed QC if $x_i > x_{\max}$, or if $x_i < x_{\min}$ and $ModelScore(y_i, x_i) < t$.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Antigen-stimulation experiments (MIRA assay)

Panel design

The multiplex identification of antigen-specific T-cell receptors (MIRA) assay was set up, performed, and analyzed as described previously (67). Briefly, 2 panels of peptides were designed and tested in the assay. The Lyme-MIRA1 panel used antigens that are known to be presented in Lyme, including tiled portions of the DbpA, OspC, OspA, BBK32, BBA52, and VlsE proteins. The Lyme-MIRA2 panel used peptides derived from antigens presented via human leukocyte antigen (HLA) class II upon *B. burgdorferi* infection. The *B. burgdorferi*-derived antigens include elongation factor Tu (WP_002657015.1), BB_0418 (WP_002658797.1), p83/100 (CAA57125.1), ABC transporter (PRR58667.1), lipoprotein LA7 (WP_002657819.1), GAPDH (AAB53930.1), chaperonin GroEL (WP_002657108.1), flagellin (WP_002661938.1), OspA (WP_010890378.1), and p66 (WP_002656762.1). The peptide tiling strategy was used across the entirety of each antigen, yielding a series of peptides, each 17 amino acids (aa) long with a 7-aa overlap between peptides.

The peptides were pooled in a combinatorial fashion as described previously (67); peptides that were overlapping or in close proximity in the viral proteome were grouped together into antigen sets. Each antigen set was then placed in a subset of 5 unique pools out of 11 total pools in the Lyme-MIRA1 panel, or 6 pools out of 12 in the Lyme-MIRA2 panel, referred to as its occupancy.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Naïve antigen-stimulation experiments

A total of 304 experiments were run with the Lyme-MIRA1 panel (all "naïve" experiments, see below) and 174 with the Lyme-MIRA2 panel. For the "naïve" experiments, CD14+ monocytes were selected from peripheral blood mononuclear cells (PBMCs; Miltenyi Biotech, Auburn, CA) and stimulated with granulocyte-macrophage colony-stimulating factor (GM-CSF) and interleukin-4 (IL-4; BioLegend, San Diego, CA) to drive dendritic cell (DC) differentiation *in vitro*. On day 3, GM-CSF, IL-4, interferon-γ (IFN- γ;BioLegend, San Diego, CA), and lipopolysaccharide (LPS; Sigma-Aldrich, St. Louis, MO, or eBioscience, Inc, San Diego, CA) were added to promote DC maturation. Also on day 3, naïve T cells were isolated from PBMCs (StemCell, Vancouver, BC, Canada) and incubated overnight with IL-7. On day 4, naïve T cells were combined with the differentiated CD14+ monocytes, IL-21 (BioLegend, San Diego, CA) and a pool of all peptides present in the panel to be used for restimulation. Cultures were supplemented with IL-7, IL-15, and IL-2 every 2 to 3 days for an additional 12 to 14 days. Cells harvested from the expansion culture were divided into a series of replicate cultures, and each was restimulated using a distinct peptide pool from the panel under investigation. After incubation at 37°C for ~20 hours, each culture was stained with antibodies (BioLegend, San Diego, CA) for sorting by flow cytometry. Cells were then washed and suspended in phosphate-buffered saline containing 2% fetal bovine serum (FBS), 1 mM EDTA, and 4,6-diamidino- 2-phenylindole (DAPI) for exclusion of non-viable cells. Cells were acquired and sorted using a FACSMelody (BD Biosciences) instrument. Sorted antigen-specific (CD4+CD137+CD145+, $CD25_{lo}$) T cells were pelleted and lysed in RLT Plus buffer (Qiagen, Germantown, MD, USA) for nucleic acid isolation.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Assignment of enhanced sequences to antigens

RNA was isolated using AllPrep DNA/RNA mini and/or micro kits, according to the manufacturer's instructions (Qiagen). RNA was then reverse transcribed to cDNA using Vilo kits (Life Technologies, Carlsbad, CA, USA), and TCRβ amplification was performed using the immunoSEQ Assay described above.

After immunosequencing, the behavior of T-cell clonotypes was examined by tracking read counts across each sorted pool. True antigen-specific clones should be specifically enriched in a unique occupancy pattern corresponding to the presence of one of the query antigens in 5 or 6 pools in the Lyme-MIRA1 and Lyme-MIRA2 panels. Methods used to assign antigen specificity to TCR clonotypes have been reported previously (67). In addition to these methods, a non-parametric Bayesian model was developed to compute the posterior probability that a given clonotype was antigen specific. This model uses the available read counts of TCRs to estimate a mean-variance relationship within a given experiment, as well as the probability that a clone will have zero read counts due to incomplete sampling of low frequency clones. Together, this model considers the observed read counts of a clonotype across all pools and estimates the posterior probability of a clone responding to all valid addresses and an additional hypothesis that a clone is activated in all pools (truly activated, but not specific to any of our query antigens). To define antigen-specific clones, we identified TCR clonotypes assigned to a query antigen from this model with a posterior probability ≥0.7.

TCR sequences from MIRA were compared to the enhanced sequence list on the basis of V-gene, J-gene, and CDR3 amino acid sequences. Any exact matches between the two lists, where

medRxiv preprint doi: https://doi.org/10.1101/2021.07.30.21261353; this version posted August 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the MIRA TCR sequence was found in at least 2 separate individuals, were considered sufficient to map the enhanced sequence to the MIRA antigen.

Enhanced sequence clustering and HLA inference

Clustering of enhanced sequences was based on TCR amino acid similarity. Specifically, two TCRs were assigned to the same cluster if they shared V-gene family (and so have similar complementary determining regions [CDRs] 1 and 2), had identical length, and differed by at most 1 amino acid in the CDR3 region. Clusters with at least 5 enhanced sequences were reported (Table 2). A sequence motif representing the CDR3 amino acid sequences assigned to each cluster was generated using WebLogo (68, 69).

To assign an enhanced sequence to a single HLA subtype, a 1-tailed FET was performed between that enhanced sequence and every HLA subtype. The enhanced sequence was assigned to the HLA subtype with the lowest *P*-value; if the lowest *P*-value was >0.001, no assignment was made. Contingency tables counted the number of individuals with/without the enhanced sequence and with/without a given HLA subtype. For HLA-DQ and HLA-DP, α/β heterodimers were treated as distinct HLA subtypes; for example, individuals with 2 α subtypes and 2 β subtypes were treated as expressing all 4 possible heterodimers. An HLA subtype was assigned to an enhanced sequence cluster if a majority (>50%) of the cluster members with an assigned HLA subtype were assigned to the same subtype.

Figure 1

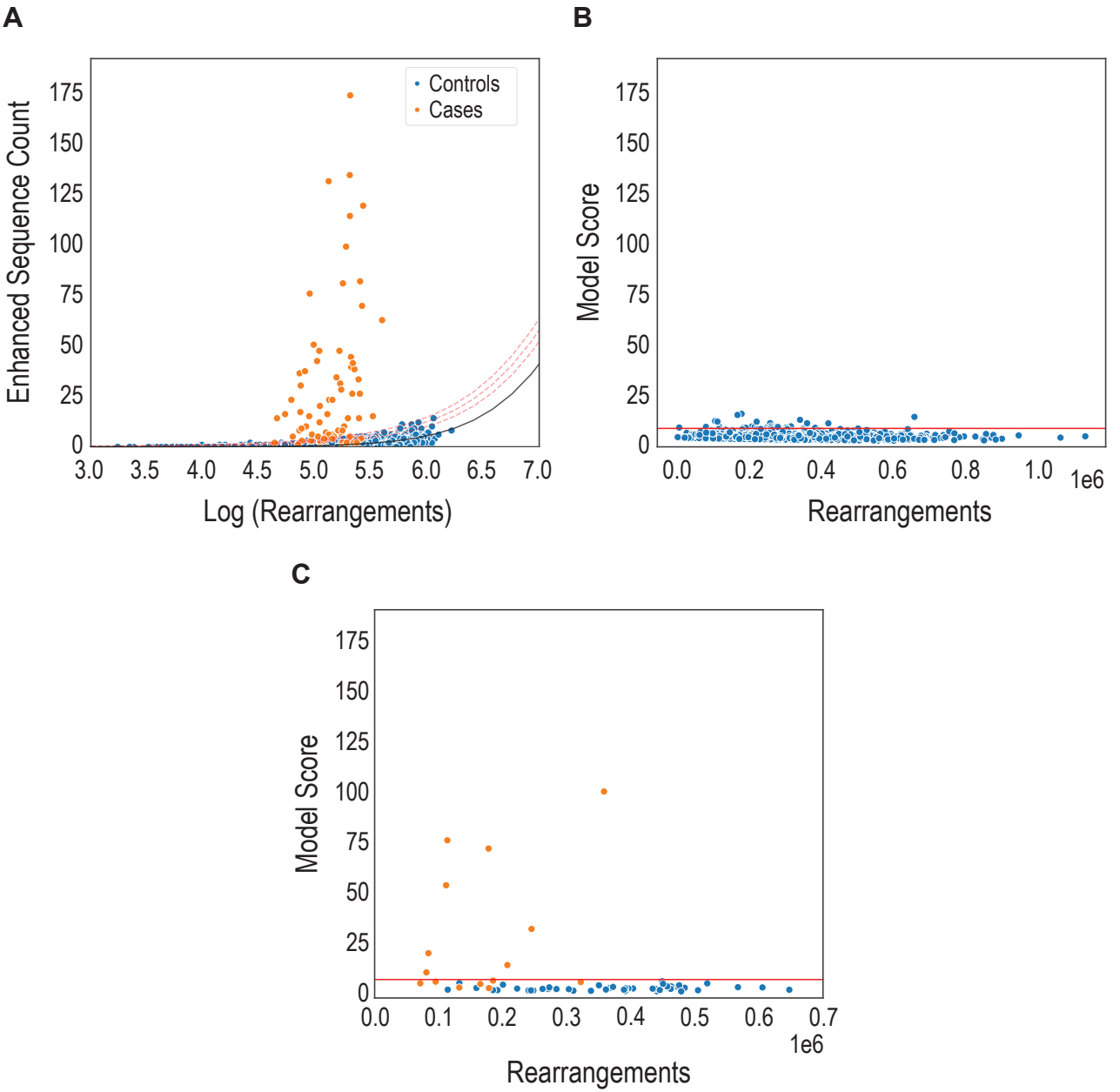

A
Controls
Cases
Enhanced Sequence Count
175
150
125
100
75
50
25
0
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
Log (Rearrangements)
B
Model Score
175
150
125
100
75
50
25
0
0.0
0.2
0.4
0.6
0.8
1.0
1e6
Rearrangements
C
Model Score
175
150
125
100
75
50
25
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
1e6
Rearrangements

Figure 2

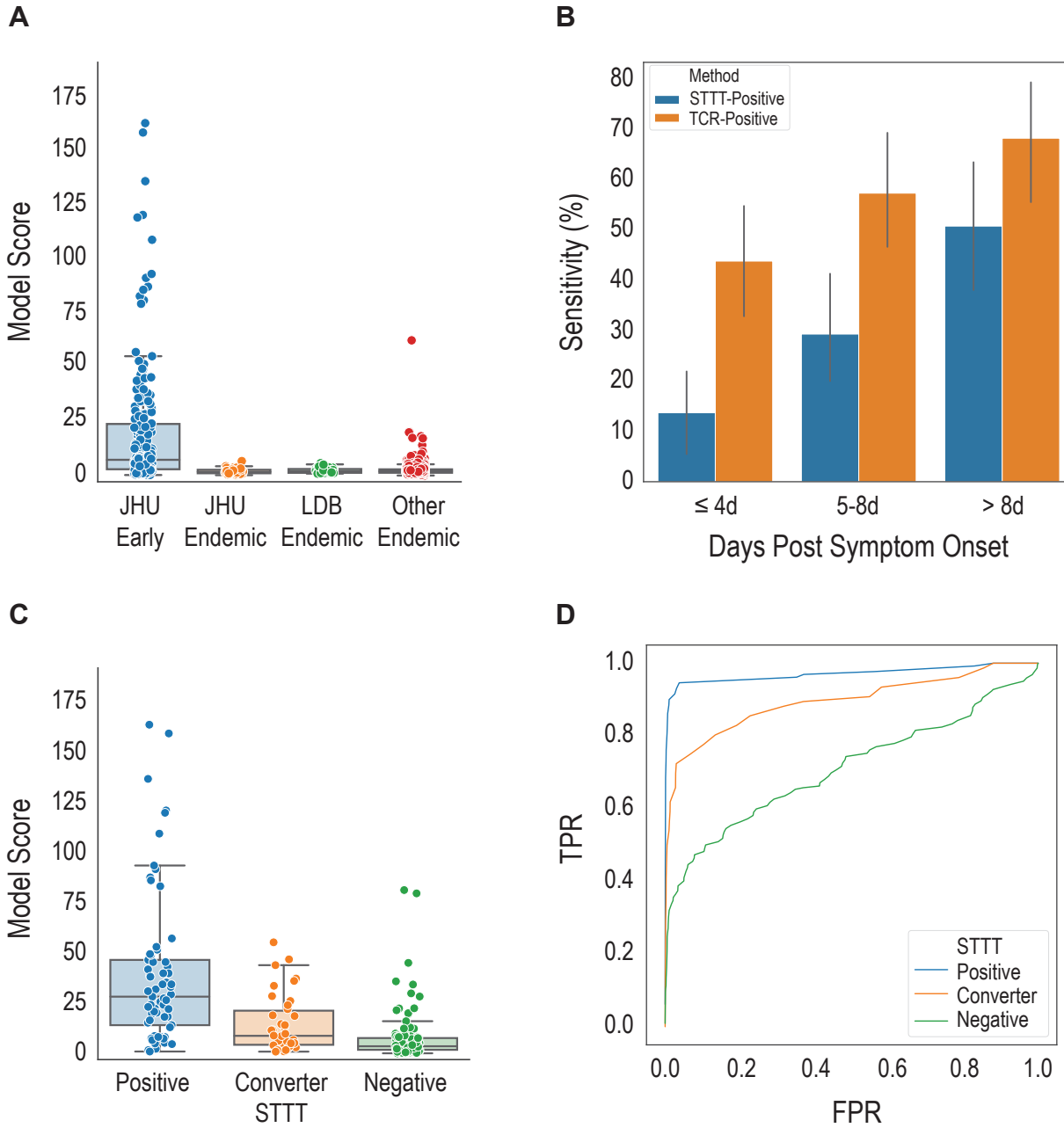

A
Model Score
175
150
125
100
75
50
25
0
JHU Early
JHU Endemic
LDB Endemic
Other Endemic
B
Method
STTT-Positive
TCR-Positive
Sensitivity (%)
80
70
60
50
40
30
20
10
0
≤ 4d
5-8d
> 8d
Days Post Symptom Onset
C
Model Score
175
150
125
100
75
50
25
0
Positive
Converter
Negative
STTT
D
TPR
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
FPR
STTT
Positive
Converter
Negative

Figure 3

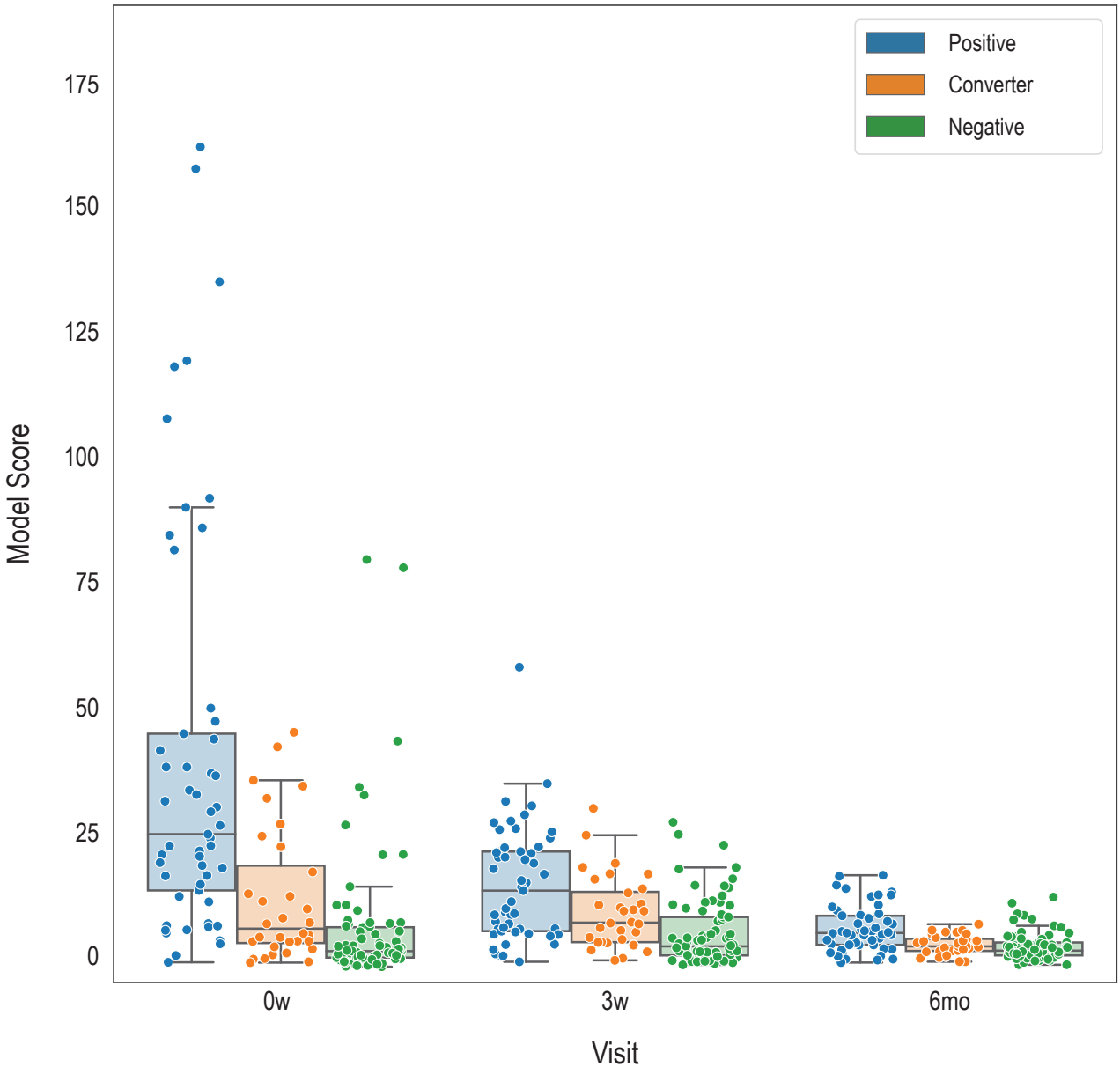

Positive
Converter
Negative
Model Score
175
150
125
100
75
50
25
0
0w
3w
6mo
Visit

Figure 4

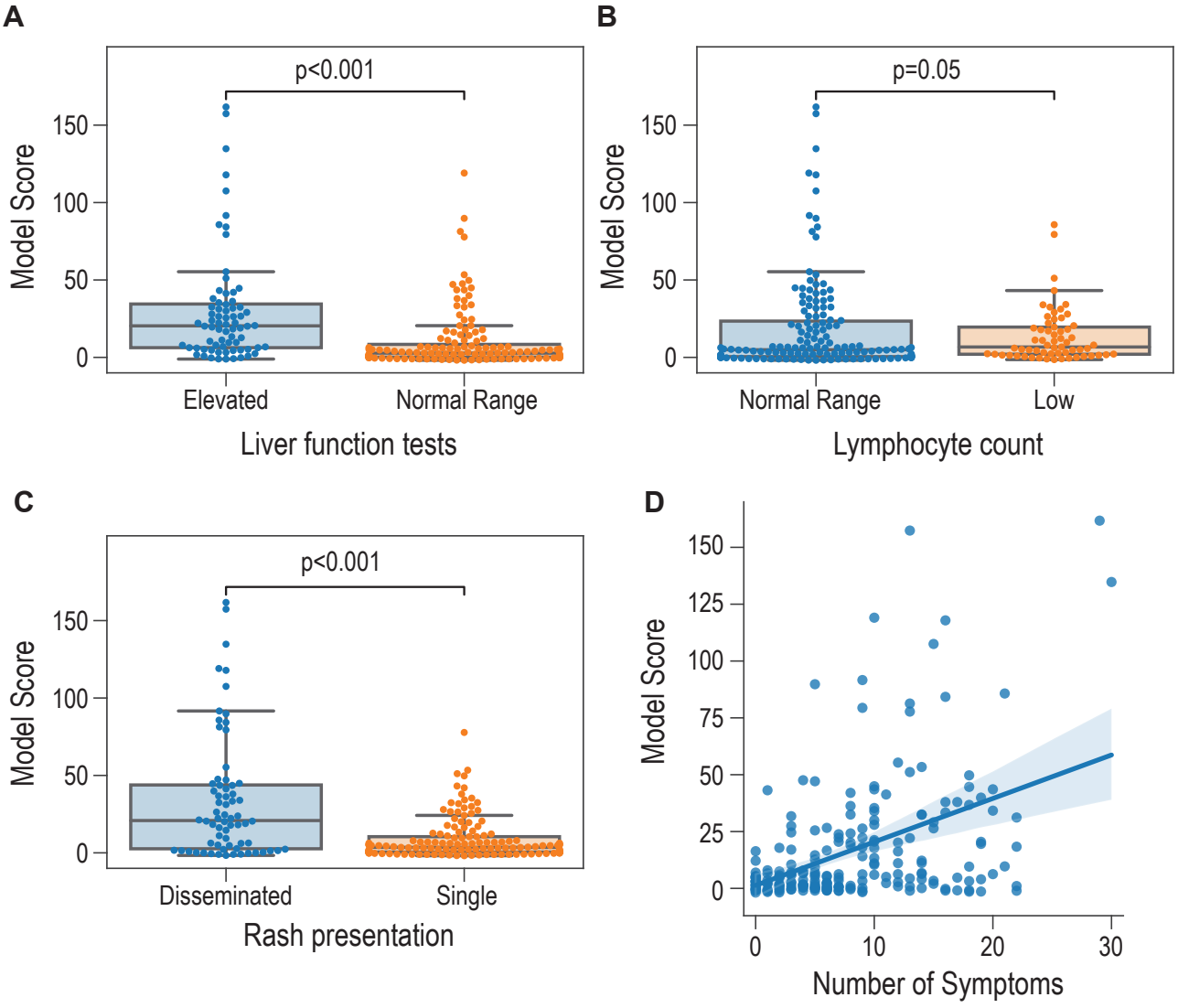

A
p<0.001
Model Score
150
100
50
0
Elevated
Normal Range
Liver function tests
B
p=0.05
Model Score
150
100
50
0
Normal Range
Low
Lymphocyte count
C
p<0.001
Model Score
150
100
50
0
Disseminated
Single
Rash presentation
D
Model Score
150
125
100
75
50
25
0
0
10
20
30
Number of Symptoms

Figure 5

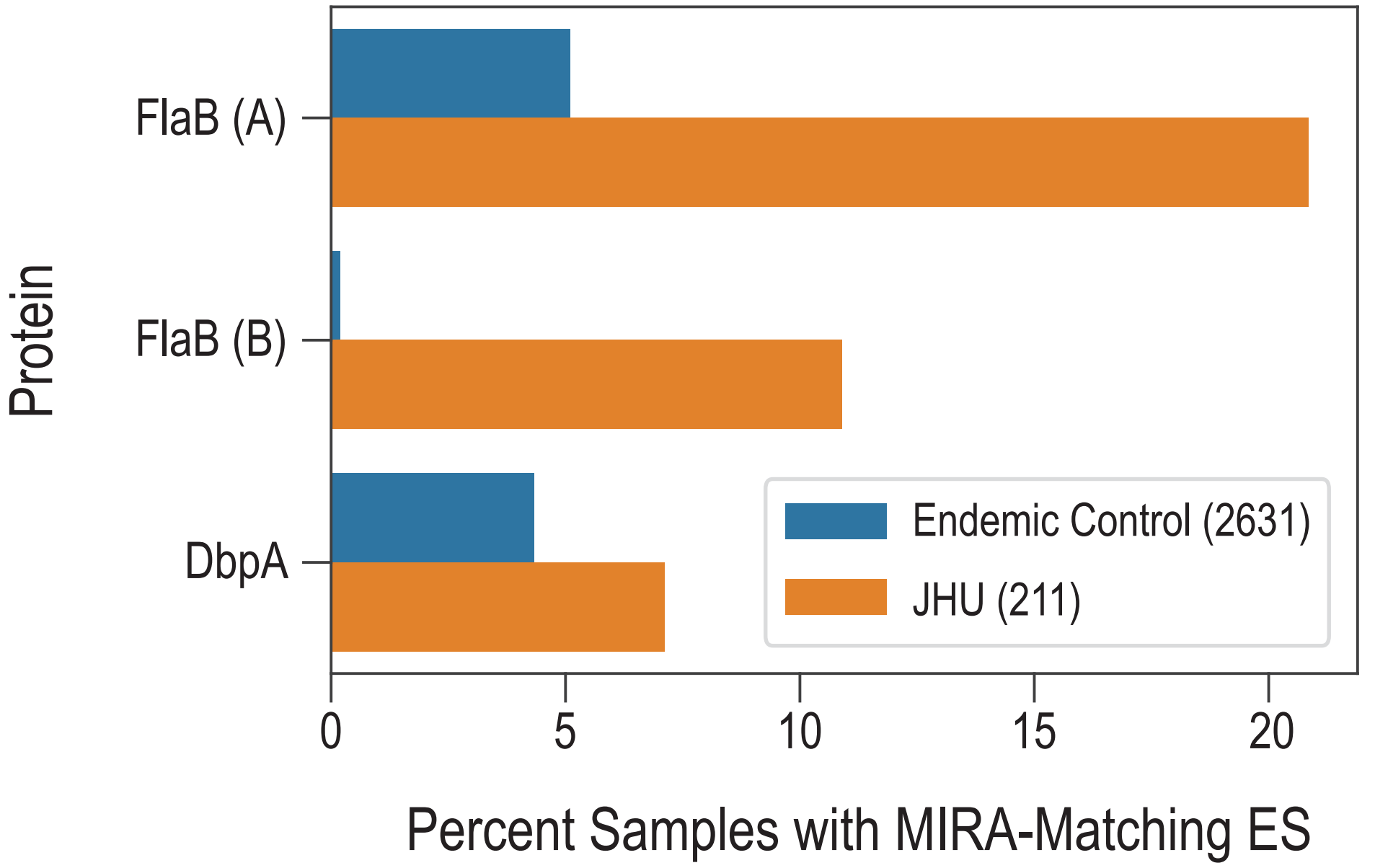
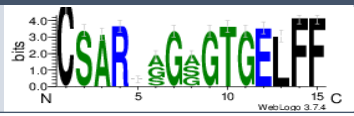
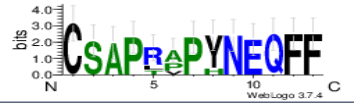

FlaB (A)
FlaB (B)
DbpA
Protein
Endemic Control (2631)
JHU (211)
0
5
10
15
20
Percent Samples with MIRA-Matching ES

Table 1. TCR classifier sensitivity stratified by serostatus.

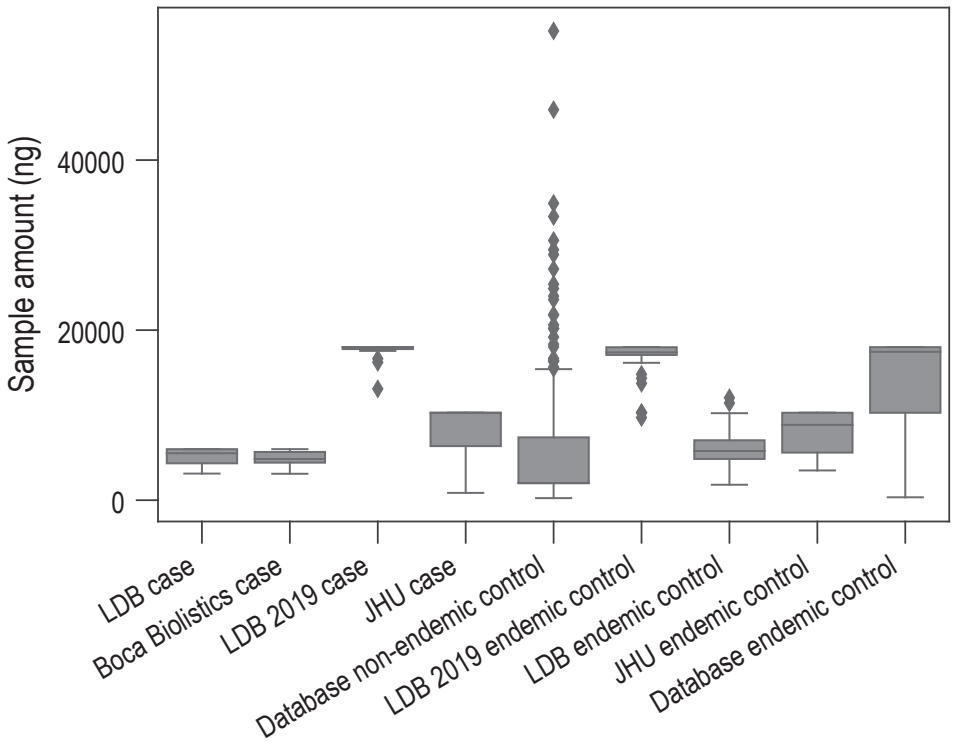
	Sensitivity	n
All JHU	0.56 (0.49–0.62)	118/211
STTT Positive	0.92 (0.84–0.98)	59/64
STTT Post-Treatment Seroconverter	0.58 (0.45–0.74)	22/38
Persistent STTT Negative	0.34 (0.27–0.43)	37/109

Table 2. TCR clusters as defined by connected components within a 1–amino-acid change in CDR3 and the same V-gene family.

Cluster	CDR3 Motif	V Family	CDR3 Length	No. Sequences	HLA	MIRA-Match TCRs
1		V20	15	65	*DPA1*01:03 + DPB1*04:01*	0
2		V06	14	12		0
3		V12	15	9	*DRB3*01:01*	0
4		V07	14	7	*DRB4*01:03*	0
5		V20	15	6	*DQA1*01:02 + DQB1*03:03*	0
6		V20	13	6	*DRB3*02:02*	6 (FlaB (A))

Supplementary Figure

A



B

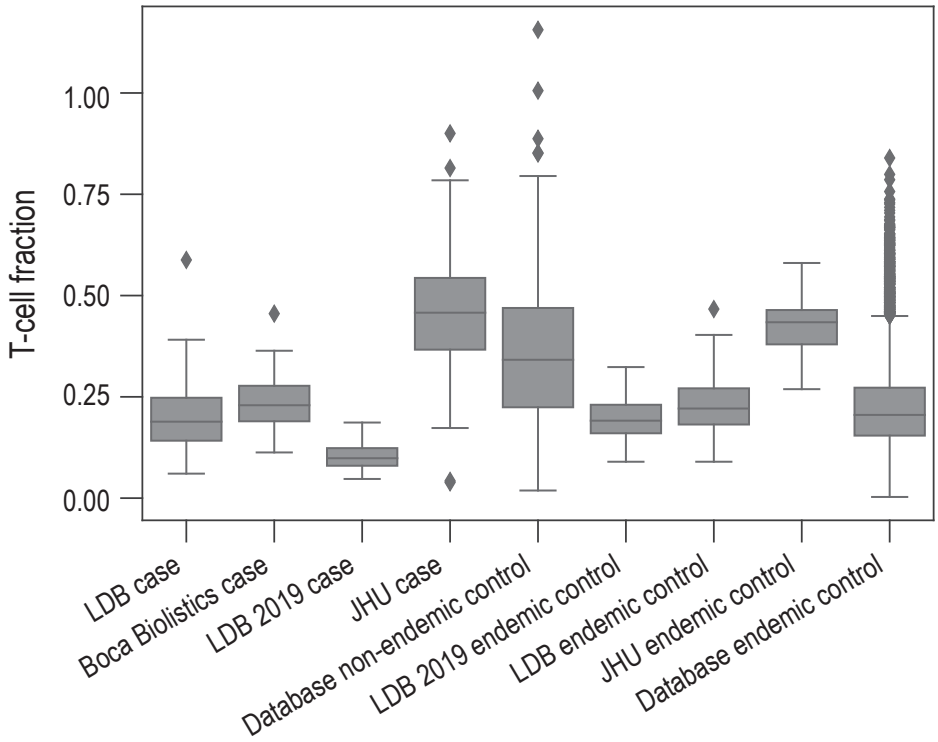


Table S1. Cohorts used for training, setting the classification threshold, and validation.

Cohort	Description	Total	Train	Threshold	Validation

LDB case	• Patients presenting prior to 2019:
	  • With EM >5 cm or an erythematous, annular, expanding skin lesion ≤5 cm, OR
	  • Without an EM/annular lesion, but with a suspected tick exposure/bite and presenting with signs or symptoms (headache, fatigue, fever, chills, or joint or muscular pain) with no history of chronic fatigue syndrome, rheumatologic disease, or multiple sclerosis
	• STTT-positive
		54	54	0	0

Boca Biolistics case	• Patients with an EM rash, positive serology results and/or evidence of a tick bite • STTT-positive	18	18	0	0
LDB 2019 case	• Patients presenting during 2019 with the same characteristics described for the LDB case subgroup, except cases were confirmed by STTT, PCR, and/or culture	15	0	0	15
JHU case	• Patients presenting with EM ≥5 cm diagnosed by a health care provider and either multiple skin lesions or ≥1 new-onset concurrent symptom • Received ≤72 hours of appropriate antibiotic treatment for early Lyme disease	211	0	0	211
Database non-endemic control	• Individuals with unknown Lyme disease status living in Lyme non-endemic regions in the US	2981	2981	0	0
LDB endemic control	• Healthy individuals living in an area of endemicity with no history of Lyme disease or tick-borne infection • STTT-negative	235	0	120	115
LDB 2019 endemic control	• Healthy individuals enrolled during 2019 with the same characteristics described for the LDB endemic control subgroup	48	0	0	48
JHU endemic control	• No clinical or serologic history of Lyme disease • STTT-negative	45	0	0	45
Database endemic control	• Individuals with unknown Lyme disease status living in Lyme-endemic regions in the US and Europe	4983	0	2507	2471
Total		**8590**	**3053**	**2627**	**2905**

Table S2. Multiple logistic regression of TCR model score on clinical features.

	Coef	SE	t	*P*>|t|	[0.025	0.975]
Intercept	-4.4736	7.323	-0.611	0.542	-18.914	9.966
STTT (positive)	17.1930	4.177	4.116	<0.001	8.956	25.430
Liver function tests (elevated)	11.7845	3.537	3.332	0.001	4.809	18.760
Single or disseminated Rashes (disseminated)	11.6126	4.020	2.888	0.004	3.685	19.541
Lymphocyte count category (normal range)	5.5743	3.559	1.566	0.119	-1.444	12.593
Sex (male)	2.0499	3.260	0.629	0.530	-4.378	8.478
Number of symptoms	1.0257	0.282	3.633	<0.001	0.469	1.582
Rash area (mm^2)	0.0080	0.012	0.644	0.521	-0.017	0.033
Days from symptom onset to sample	-0.1968	0.184	-1.067	0.287	-0.561	0.167
Age	-0.0866	0.102	-0.847	0.398	-0.288	0.115

Table S3. Counts of enhanced sequences mapped to each protein by MIRA.

Protein (Antigen)	No. Enhanced Sequences	Total No. Matches Across Experiments
FlaB (A)	7	99
FlaB (B)	1	5
DbpA	1	4
Total	**9**	**108**