

1 **Immunosequencing of the T-Cell Receptor Repertoire Reveals Signatures**
2 **Specific for Identification and Characterization of Early Lyme Disease**

3
4 **Authors:** Julia Greissl^a, Mitch Pesesky^b, Sudeb C. Dalai^{b,c*}, Alison W. Rebman^d, Mark J.
5 Soloski^d, Elizabeth J. Horn^e, Jennifer N. Dines^{b*}, Darcy B. Gill^b, Rachel M. Gittelman^{b*},
6 Thomas M. Snyder^b, Ryan O. Emerson^{b*}, Edward Meeds^a, Thomas Manley^{b*}, Ian M. Kaplan^{b*},
7 Lance Baldo^{b*}, Jonathan M. Carlson^a, Harlan S. Robins^b, John N. Aucott^{d#}

8
9 **Affiliations:**

10 ^aMicrosoft Research; Redmond, Washington, USA, and Cambridge, UK.

11 ^bAdaptive Biotechnologies; Seattle, Washington, USA.

12 ^cStanford University School of Medicine; Stanford, California, USA.

13 ^dLyme Disease Research Center, Division of Rheumatology, Department of Medicine, Johns
14 Hopkins University School of Medicine; Baltimore, Maryland, USA.

15 ^eLyme Disease Biobank; Portland, Oregon, USA.

16 #Address correspondence to John N. Aucott, jaucott2@jhmi.edu

17 Julia Greissl and Mitch Pesesky contributed equally to this work. Jonathan M. Carlson, Harlan S.
18 Robins, and John N. Aucott contributed equally to the oversight of this study. Author order was
19 determined by mutual agreement.

20 *Author was employed by Adaptive Biotechnologies at time of research. Present address:

21 Sudeb C. Dalai, Palo Alto Medical Foundation, Palo Alto, California, USA.

1 Jennifer N. Dines, Scipher Medicine, Waltham, MA, USA.

2 Rachel M. Gittelman, Guardant Health, Inc., Redwood City, California, USA.

3 Ryan O. Emerson, A-Alpha Bio, Seattle, Washington, USA.

4 Thomas Manley, IGM Biosciences, Inc., Mountain View, California, USA.

5 Ian M. Kaplan, Merck Research Labs, South San Francisco, California, USA.

6 Lance Baldo, Freenome, Inc., South San Francisco, California, USA.

7

8 **Running title:** T-Cell Receptor Immunosequencing in Lyme Disease

9 **Keywords:** *Borrelia burgdorferi*, Lyme disease, T-cell receptor, immunosequencing, standard
10 two-tiered testing (STTT), diagnostic

11

12 The authors declare a conflict of interest. M.P., J.N.D., D.B.G., R.M.G., T.M.S., and I.M.K.

13 declare current or former employment with and equity ownership in Adaptive Biotechnologies.

14 S.C.D. declares former employment with Adaptive Biotechnologies, employment with Stanford

15 University School of Medicine, and equity ownership in Adaptive Biotechnologies. A.W.R.

16 declares institutional support from the Steven and Alexandra Cohen Foundation and Global

17 Lyme Alliance. M.J.S. declares institutional support from the Steven and Alexandra Cohen

18 Foundation and National Institutes of Health grant P30 AR070254. E.J.H. declares compensation

19 from the Lyme Disease Biobank and institutional support from Bay Area Lyme Foundation,

20 Steven and Alexandra Cohen Foundation, and Adaptive Biotechnologies. R.O.E. and T.M.

21 declare employment with Adaptive Biotechnologies during the time of this research. L.B. and

22 H.S.R. declare employment with, equity ownership in, and leadership for Adaptive

23 Biotechnologies. J.N.A. declares consulting with Tarsus Pharmaceuticals and Pfizer;

1 participation in an advisory board with Adaptive Biotechnologies; expert testimony; membership
2 in the Bay Area Lyme Foundation Scientific Advisory Board; Past Chair, 2018, HHS Tick-borne
3 Disease Working Group, Office of HIV/AIDS and Infectious Disease Policy, Office of the
4 Assistant Secretary of Health, Department of Health and Human Services; and institutional
5 support from the Steven and Alexandra Cohen Foundation and Global Lyme Alliance. J.G.,
6 E.M., and J.M.C. declare employment with and equity ownership in Microsoft.

7

8

9

10

11

12

13

14

15

16

17

18

19

20

1 **ABSTRACT**

2 Highly specific T-cell responses play key roles in pathogen clearance and maintaining
3 immunologic memory. Next-generation sequencing of the T-cell receptor (TCR) repertoire is an
4 emerging diagnostic technology that capitalizes on the specificity of T-cell responses to probe
5 pathogen exposure. The spirochete *Borrelia burgdorferi* (*Bb*) wields an array of antigens with
6 dynamic and complex immunogenic potential, and application of TCR immunosequencing to
7 characterize *Bb* infection presents opportunities to improve detection of early Lyme disease
8 (LD). By immunosequencing TCR repertoires in blood samples from 3 independent cohorts of
9 patients with early LD and controls from Lyme-endemic/non-endemic regions, we identified 251
10 public, LD-associated TCRs. These TCRs were used to train a classifier for detection of early
11 LD. The classifier identified LD with 99% specificity and showed 1.9-fold higher sensitivity
12 (56% vs 30%) compared with standard two-tiered testing (STTT). TCR positivity predicted
13 subsequent seroconversion in 37% of STTT-negative patients, suggesting that the T-cell
14 response is detectable before the humoral response. Higher TCR scores were associated with
15 clinical measures of disease severity, including abnormal liver function tests, disseminated rash,
16 and number of symptoms. A subset of LD-associated TCRs mapped to *Bb* antigens, supporting
17 specificity of this approach. These results suggest that TCR testing may be a highly specific and
18 sensitive approach for identifying LD, particularly in the initial days of illness.

19

1 INTRODUCTION

2 Lyme disease (LD) is the most common tick-borne illness in the United States, with more than
3 450,000 estimated new cases annually (1–3). In the United States, LD is initiated by infection
4 with the spirochetal bacterium *Borrelia burgdorferi* (*Bb*) transmitted from infected *Ixodes* ticks
5 (1). In the days to weeks after the initial tick bite, symptoms may include a characteristic
6 erythema migrans (EM) rash and nonspecific flu-like symptoms while signs and symptoms of
7 disseminated infection can affect the joints, nervous system, heart, or other areas of the skin (4).
8 Presence of EM rash in Lyme-endemic areas is highly suspicious for LD and warrants immediate
9 treatment without further testing. However, guidelines recommend serologic testing to support a
10 diagnosis of LD for individuals with absent or atypical EM or suspected disseminated infection
11 (5, 6). The most common laboratory test for LD is standard two-tiered testing (STTT), which
12 combines an initial enzyme-linked immunosorbent assay (ELISA) with a more specific
13 immunoblot for positive or equivocal samples to detect antibodies against *Bb* (7, 8). However,
14 STTT has some notable limitations, including poor sensitivity (25-50%) in the very early stages
15 of infection, when most serologic testing is conducted (9, 10). This is likely attributable to the
16 kinetics of the humoral response, as sensitivity of STTT improves with time, potentially
17 exceeding 99% in untreated individuals with later-stage disease (11). Up to 60% of individuals
18 testing negative within the first days to weeks after onset of symptoms may test positive upon re-
19 testing 30 days later (10, 12–17). Additionally, false positives and interlaboratory variability may
20 occur with STTT due to poor specificity and weak immunoreactivity (18, 19). These limitations
21 underscore the unmet clinical need for improved methods for diagnosing LD, especially in the
22 early stages of infection.

1 Infection with *Bb* elicits a T-cell response with kinetics that differ from the humoral response
2 (20–22). Evaluation of cytokine/chemokine profiles suggests that an active T-cell response is
3 induced during acute infection, even in the absence of seroconversion, and returns to normal
4 levels after treatment and symptom resolution (23). Thus, interrogating the T-cell response may
5 be a useful strategy for detecting LD during the earliest stages of illness and understanding the
6 immune response throughout disease progression.

7 T-cell responses rely on the capacity of unique T-cell receptors (TCRs) to recognize specific
8 peptide antigens presented on the cell surface by human leukocyte antigen (HLA) proteins for
9 antigen-induced clonal expansion and differentiation into effector cells. Some of these expanded
10 T cells become part of the memory compartment, where they can reside for many years as clonal
11 populations of cells with identical TCR rearrangements (24–26). While the diversity of TCR
12 recombination means that most of these disease-specific TCRs are “private,” or highly unique to
13 one individual, part of the T-cell response to a disease is “public,” with identical amino acid
14 sequences observed across multiple individuals, particularly those with shared HLA backgrounds
15 (27). The public, disease-associated TCRs can be identified using a case/control study design
16 (28–31) and matched to specific antigens through multiplex identification of antigen-specific
17 TCRs (MIRA) (28, 32). Because these public clones are antigen- and HLA-specific, enrichment
18 of such clones serves as a signature of infection in a given HLA context (28, 29). This approach
19 has been successful in identification of past infection by cytomegalovirus (CMV) (28) and severe
20 acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (29, 30, 33).

21 Given the well-characterized roles of CD8⁺ and CD4⁺ T cells in viral control and clearance (34),
22 it is not surprising that strong T-cell responses are elicited by SARS-CoV-2 and mediated by
23 similar antigens and TCRs across individuals (29–31). While CD4⁺ T cells are known to

1 differentiate into effector cell types with roles in promoting cellular and humoral immunity
2 following many bacterial infections (35), the publicity of T-cell responses to bacterial pathogens
3 has yet to be characterized. To better understand the utility of TCR immunosequencing in the
4 context of bacterial infection, we leveraged the previously described case-control study design
5 approach (28–31) to characterize LD-specific TCRs and develop a classifier to aid in diagnosis
6 of LD. We tested the performance of the classifier in case and control cohorts distinct from those
7 used in classifier training and assessed the correlation between T-cell responses and clinical
8 features of LD. We also mapped a subset of the identified LD-associated TCRs to specific *Bb*
9 antigens, supporting the biologic specificity of the TCR immunosequencing approach.

10

11 **RESULTS**

12 **Identification of LD-associated TCRs**

13 To characterize the T-cell response to LD, we immunosequenced TCR β in samples from 3 LD
14 cohorts and a database of controls from Lyme-endemic and non-endemic regions (Tables 1 and
15 2). Public TCRs associated with early LD were identified from a subset of these samples
16 comprising a case/control training dataset of 72 patients identified from the LDB (n=54) and
17 Boca (n=18) cohorts who presented with STTT-positive early LD prior to 2019 and control
18 repertoires (n=2,981) from a database of healthy individuals from non-endemic regions recruited
19 for other studies and presumed to be LD-negative (Figure 1A). A total of 251 public TCRs
20 associated with early LD, referred to as “enhanced sequences” (ESs), were identified based on
21 statistical enrichment in cases compared with controls.

22

1 **Enhanced TCR sequences are highly specific for identifying early LD**

2 Clonal expansion in response to *Bb* infection should lead to enrichment of LD-associated ESs in
3 patients with LD. A low-level presence of LD-associated ESs would also be expected among
4 healthy individuals. The observed background rate of ESs is a function of the number of
5 sequenced T cells and the number of unique TCRs in each sample. Empirically, the background
6 number of ESs in our training population fit a logistic-growth curve. To leverage the number of
7 ESs and total unique productive TCR rearrangements as a diagnostic classifier, we modeled the
8 number of ESs as a logistic-growth function of the number of unique productive TCRs sampled
9 from a repertoire and fit this model to the 2,981 control repertoires in the training data (Figure
10 1A, right graph; black line represents the model fit). The resulting model provides a normalized
11 estimate of the degree to which the ES signature of a case sample deviates from what is typically
12 seen in our control data, expressed in standard deviations from the mean (Figure 1A, right graph,
13 red dashed lines). This approach carefully controls specificity by considering thousands of
14 control repertoires. The final positive/negative call threshold was set to a specificity of 99% on
15 an independent set of controls from Lyme-endemic areas (n=2,627, consisting of 2,507 presumed
16 LD-negative samples from our database and 120 STTT-negative controls from Lyme-endemic
17 areas from the LDB cohort [Figure 1B; Tables 1 and 2]).

18 To further verify the classifier and call threshold among patients with demographics similar to
19 the training dataset, we applied the resulting model to a holdout set of samples from the LDB
20 cohort collected in 2019 that included 15 laboratory-confirmed-positive (by STTT and/or
21 polymerase chain reaction [PCR]) cases of early LD and 48 laboratory-confirmed-negative (by
22 STTT) controls from Lyme-endemic areas with no history of Lyme or tick-borne infection

1 (Figure 1C; Tables 1 and 2). Overall, 8 of 15 (53%) early LD samples and 0 of 48 (0%) controls
2 from Lyme-endemic samples were identified as TCR-positive by the classifier.

3

4 **Sensitivity of TCR repertoire analysis identifying early LD**

5 To further assess the generalizability of our approach in an independent cohort collected with
6 different protocols, we evaluated performance of the TCR classifier using samples from STTT-
7 positive and STTT-negative patients who had clinically diagnosed early LD and documented EM
8 at the time of their enrollment in the JHU cohort (Tables 1 and 2). Consistent with results from
9 the LDB and Boca cohorts, 118 of 211 (56%) patients diagnosed with early LD were classified
10 as TCR-positive (Table 3), while only 32 of 2,631 (1.2%) control samples from Lyme-endemic
11 areas were TCR-positive (0/115 from LDB, 1/45 from JHU, and 31/2,471 from a database of
12 individuals with unknown LD status living in Lyme-endemic regions in the United States and
13 Europe) (Figure 2A). In contrast, only 64 of 211 (30%) LD-positive JHU samples were STTT-
14 positive, representing a 1.9-fold reduction in sensitivity relative to TCR immunosequencing. To
15 further confirm the specificity of the TCR signature, we evaluated 21 samples from individuals
16 with PCR-confirmed anaplasmosis (n=4) or babesiosis (n=17) from the LDB and DLS cohorts.
17 All 21 samples were negative according to the classifier (Figure 2A).

18 When examined as a function of self-reported duration of symptoms, sensitivity of T-cell and
19 antibody tests increased with time, yet T-cell testing showed greater sensitivity at early time
20 points (Figure 2B). If T-cell responses typically precede and facilitate B-cell responses in
21 response to *Bb* infection, consistent with immune response to viral and bacterial pathogens (36,
22 37), then TCR positivity among a seronegative cohort should predict subsequent seroconversion,
23 and a majority of STTT-positive samples should also be TCR-positive. Indeed, among the JHU

1 case cohort, TCR testing was positive in 59 of 64 (92%) STTT-positive cases (58 of 61 [95%]
2 STTT-positive by IgM), compared with 59 of 147 (40%) STTT-negative cases (Table 3),
3 indicating that a detectable antibody response is highly predictive of a detectable T-cell response.
4 Importantly, of the 59 TCR-positive/STTT-negative individuals, 22 (37%) subsequently
5 seroconverted from STTT-negative at study enrollment to STTT-positive at the first post-
6 treatment follow-up visit (~3 weeks after enrollment), while only 16 of 88 (18%) individuals
7 who were TCR-negative at baseline seroconverted over the same time period ($P=0.01$, Fisher's
8 exact test). Stratification of the JHU cohort demonstrated that median TCR model scores (Figure
9 2C) and classifier sensitivity (Figure S3 Table 3) were highest among individuals who were
10 STTT-positive at enrollment, intermediate among those who seroconverted post treatment, and
11 lowest among individuals who remained persistently STTT-negative, highlighting differences in
12 the extent of LD-associated T-cell expansion in these serologically defined subpopulations.
13 Taken together, these data indicate that detectable LD-associated T cells typically expand prior
14 to detectable antibodies, suggesting that identification of LD may be aided by TCR
15 immunosequencing during early phases of *Bb* infection.

16

17 **Disease-associated TCRs and seropositivity wane after treatment**

18 Previous data suggest that T-cell and humoral immune responses exhibit differing dynamics over
19 the course of *Bb* infection (21). To probe these dynamics, we evaluated TCR repertoires in
20 longitudinal samples from individuals enrolled in the JHU cohort. Patients were either antibiotic-
21 naïve or had initiated 3 weeks of oral doxycycline treatment within the 72 hours prior to study
22 enrollment, with samples collected at enrollment, immediately after treatment, and at 6 months
23 post treatment. Immunosequencing of samples from 161 patients with samples available at all

1 time points revealed that TCR responses waned significantly in the 6 months following treatment
2 (Figure 2D) (30). Median model scores decreased from 6.1 to 2.5, and model sensitivity
3 decreased from 56% (91/161) at enrollment to 32% (51/161) 6 months post treatment. Notably,
4 the sensitivity of STTT also declined over the same time period, from 33% at enrollment to 12%
5 at 6 months post treatment (14 of 115 patients with available STTT results) (Figure 2E). Similar
6 to the results shown in Figure 2C, TCR model scores were higher across all time points among
7 individuals who were STTT-positive at baseline compared with those who were STTT-negative
8 (Figure 2D). T-cell testing was more sensitive than STTT for identification of LD post treatment
9 (32% vs 12% at 6 months post treatment (Figure 2E) including in patients who did not undergo
10 IgG seroconversion (not shown).

11

12 **T-cell responses correlate with clinical measures of LD severity**

13 The strong correlation observed between antibody and T-cell responses highlights the
14 interconnectedness of the immune response in early LD and may also reflect underlying
15 pathogen burden, disease severity, or other clinical measures that drive the immune response. We
16 therefore explored potential associations between clinical parameters previously reported in the
17 JHU study (38) and the strength of the T-cell response as measured by the TCR model score at
18 diagnosis. In both univariate analyses (Figure 3A-D) and a multiple regression model (Table S1)
19 that adjusted for sex, age, and serostatus, higher TCR scores were associated with markers of
20 disease severity, including ≥ 1 elevated liver function test (aspartate aminotransferase, alanine
21 aminotransferase, and alkaline phosphatase), disseminated rash, and the number of pretreatment
22 LD-associated symptoms. Notably, the highest model scores were observed among STTT-
23 positive individuals with disseminated rash and elevated liver function tests (Figure 3A, 3C, S2).

1 Sex, age, size of rash, and an abnormally low lymphocyte count were not associated with a
2 difference in TCR model scores in this cohort (Table S1).

3

4 **Antigen specificity of Lyme-associated TCRs**

5 Analysis of TCR sequence similarity clustered 105 of the 251 ESs (42%) to one of 6 clusters, 5
6 of which were statistically associated with an HLA class-II heterodimer (Table 4 and
7 Supplementary Materials). In a MIRA experiment that queried T cells derived from 395 healthy
8 donors against 777 peptides from 26 *Bb* proteins, 6 ESs from *HLA-DRB3*02:02*-associated
9 cluster 6 exactly matched TCRs that were mapped to the flagellin B (FlaB)-derived peptide
10 MIINHNTSAINASRNNG, providing direct evidence of *Bb* specificity for these ESs. One of
11 these TCR sequences was found in 26 individuals, all of whom expressed *HLA-DRB3*02:02*
12 (among the 25 with available typing). A Basic Local Alignment Search Tool (BLAST) analysis
13 of this peptide showed limited sequence similarity with non-*Borrelia* pathogens, and the
14 associated TCRs were highly enriched in JHU cases compared to endemic controls (Fig. 4).
15 Three additional non-clustered TCRs were mapped to 3 separate antigens, 2 in FlaB and 1 in
16 Dbpa, and each of these were similarly enriched in cases compared to controls (Fig. 4). As a
17 negative statistical control, none of the LD ESs matched any TCRs derived from 507 individuals
18 that were previously queried against 325 SARS-CoV-2-derived peptides.

19

20 **DISCUSSION**

21 Results from this study provide proof of principle that the high-throughput TCR sequencing and
22 machine learning approach we previously applied for identification of viral infections (eg, CMV

1 and SARS-CoV-2 [28, 29, 30]) can be adapted for identification of *Bb* infection. We demonstrate
2 that TCR repertoire characterization for sequence-based identification of public, disease-specific
3 TCRs in the setting of acute bacterial infection is a powerful and generalizable approach to aid in
4 diagnosing disease.

5 Identification of 251 LD-associated enhanced TCR sequences served as the basis for training a
6 classifier capable of sensitive and specific detection of LD across 3 independent cohorts of
7 patients with laboratory-confirmed and/or clinically diagnosed early LD. The TCR classifier
8 identified patients with early LD with 1.9-fold greater sensitivity than STTT (56% vs 30%),
9 while maintaining a specificity of 99%. Enhanced sensitivity was most apparent in early illness
10 (44% vs 14%, or 3.1-fold increase in sensitivity ≤ 4 days since symptom onset), and TCR
11 positivity was predictive of subsequent STTT seroconversion in 37% of STTT-negative
12 individuals, consistent with expansion of LD-specific T cells preceding detectable antibody
13 responses.

14 TCR positivity was associated with STTT positivity at enrollment (92%) and did not decline as
15 rapidly as serologic responses following treatment. Longitudinal analyses showed that TCR
16 scores decreased with time post treatment, consistent with diminishment of the T-cell response
17 with resolution of disease, yet remained more sensitive than STTT for identification of LD post
18 treatment, including in patients who did not undergo IgG seroconversion. This observation
19 suggests that TCR testing may be able to identify LD even in the absence of seroconversion at
20 convalescence, which has been observed among individuals treated early in the course of disease
21 (39). Furthermore, higher TCR scores correlated with clinical measures of disease, including
22 elevated liver function tests, disseminated rash, and number of disease-associated symptoms.
23 This suggests that the magnitude of the T-cell response is associated with the degree of disease

1 severity. Finally, we mapped a subset of Lyme-associated ESs to known *Bb* antigens, supporting
2 the high biologic specificity of a TCR immunosequencing approach (40).

3 Our data imply that T-cell activation precedes the humoral response, although both T-cell–
4 dependent and –independent responses have been implicated in clearance of *Borrelia* infection
5 (37, 41). Currently, the primary CDC-recommended testing strategy for LD is STTT, which
6 probes the humoral response to *Bb* (7, 8). Given the high prevalence of testing performed in
7 patients during the early stages of LD infection, when sensitivity of STTT is poor (9), alternative
8 approaches to detecting *Bb* are needed. Recently, modified two-tiered testing (MTTT) algorithms
9 have been adopted that utilize 2 sequential ELISAs, eliminate immunoblotting, and demonstrate
10 improved sensitivity over STTT in early LD (42). Even so, almost half of individuals with PCR-
11 confirmed *Bb* infection do not produce a detectable serologic response (43). Our results confirm
12 the ability of the TCR assay to identify LD in a large proportion of STTT-negative individuals
13 prior to seroconversion. These data support further studies directly comparing TCR testing to
14 MTTT to better understand the potential of TCR testing as both an alternative and
15 complementary diagnostic approach to any serologic testing modality.

16 The present analysis is limited to retrospective evaluation of samples previously collected from
17 well-defined, prospective cohorts of clinically confirmed and/or laboratory-confirmed early LD.
18 Additional prospective clinical validation studies are needed to further characterize the
19 advantages of TCR testing relative to serology in scenarios where the spectrum of presenting
20 illness, symptomology, and duration may vary. For example, further studies are needed to
21 understand the utility of TCR testing in patients with persistent symptoms in the setting of LD.
22 Approximately 10% to 20% of patients treated for LD experience long-term symptoms lasting 6
23 months or more after treatment, known as post-treatment Lyme disease (PTLD) (6, 44). TCR

1 testing may provide a means for diagnosing these patients and minimize the potential for
2 misleading interpretation of positive results (23, 45). In addition, we found that 11 of 12
3 individuals who presented with EM rash within the JHU cohort, but were both STTT-negative
4 and *Bb* PCR-negative, were also TCR-negative. These data may reflect sensitivity limitations of
5 both immune assays but could also be attributable to difficulties in clinically discriminating
6 between LD and other similar tick-borne illnesses, such as STARI (43), suggesting exploration
7 of TCR testing for differential diagnosis of LD and other similar tick-borne illnesses as another
8 important area of future investigation. Although we examined potential assay cross-reactivity
9 against *Anaplasma* and *Babesia* (Figure 2A), evaluation of cross-reactivity with other pathogens
10 should be performed. Nevertheless, it is noteworthy that antigen mapping indicates that a subset
11 of the identified TCRs are highly specific for known *Bb* proteins, including FlaB, a protein
12 known to bind cross-reactive antibodies in the immunoblotting component of STTT (46).

13 Results of this study suggest that TCR testing can have high clinical utility as a sensitive and
14 specific diagnostic for LD that may facilitate earlier diagnosis of LD and initiation of antibiotic
15 treatment to prevent development of severe illness in patients lacking definitive clinical
16 signs/symptoms. Application of the present TCR classifier as a diagnostic assay will be
17 prospectively evaluated relative to 2-tiered testing in an ongoing clinical validation study in
18 patients presenting with suspected LD (NCT04422314).

19

20 **MATERIALS AND METHODS**

21 **Study design**

1 This study employed previously collected samples to develop and validate a model for
2 identifying early LD with high sensitivity and specificity. TCR repertoires were sequenced from
3 8,585 participants, including 298 cases of LD enrolled in prospective studies and 8,287 controls
4 obtained from a variety of sources. Allocation of samples to sets used for classifier training,
5 setting the classification threshold, and validation of the TCR assay was prespecified; all
6 available samples meeting the quality control (QC) and assignment criteria were included in the
7 analysis (see Assignment of Cohorts and Tables 1 and 2). The primary endpoint of the study was
8 evaluation of sensitivity in the JHU cohort, which was selected based on the conservative
9 enrollment criteria for that cohort. Inclusion and exclusion criteria for all cohorts are detailed
10 below.

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

18

19 **Study cohorts**

20 For a detailed summary of all cohorts, see Tables 1 and 2.

21

1 *LDB cohort*

2 LDB is a program of the Bay Area Lyme Foundation. The LDB cohort enrolled individuals from
3 the East coast and upper Midwest regions of the United States who presented with signs or
4 symptoms consistent with early LD. Included patients presented with EM rash greater than 5 cm
5 in diameter or an erythematous, annular, expanding skin lesion of 5 cm or less or presented with
6 signs or symptoms (headache, fatigue, fever, chills, or joint or muscular pain) without an
7 EM/annular lesion, but with a suspected tick exposure or tick bite, and with no history of chronic
8 fatigue syndrome, rheumatologic disease, or multiple sclerosis. Individuals with tick-bite
9 reactions (eg, a non-annular erythematous macule at the site of the tick bite) without EM or
10 expanding annular lesion were excluded, as were those who had initiated antibiotics more than
11 48 hours before enrollment. Healthy individuals living in the same regions with no history of LD
12 or tick-borne infection were enrolled as controls from Lyme-endemic areas. Real-time PCR for
13 *Bb* detection in whole blood and serologic testing for antibodies against *Bb* using the
14 immunoassays comprising STTT (non-reflexive) were conducted for all individuals.
15 Laboratory-confirmed early LD samples were defined as being STTT-positive, PCR-positive
16 (sample or culture), or having 2 positive ELISAs and an EM of 5 cm or larger in diameter.
17 Laboratory-confirmed controls from Lyme-endemic areas were defined as STTT-negative.
18 Additional details and baseline clinical characteristics of this cohort were previously published
19 (47). Only laboratory-confirmed samples that were STTT-positive were used in model training.
20 Only laboratory-confirmed-positive (by STTT and/or PCR) cases and laboratory-confirmed-
21 negative (by STTT) controls from Lyme-endemic areas were included in analyses of model
22 performance or to establish the final call threshold. Nine cross-reactivity case samples that were

1 positive by PCR for *Anaplasma* (n=4) or *Babesia* (n=5) were used for additional specificity
2 testing.

3

4 *Discovery Life Sciences (DLS) cohort*

5 Blood samples from 12 individuals found to be *Babesia*-positive by PCR testing were acquired
6 from Discovery Life Sciences for use in specificity testing.

7

8 *Boca cohort*

9 Specimens were collected from antibiotic-treatment-naïve patients recruited at clinical sites
10 throughout New York and New Jersey who presented with acute symptomology of a tick-borne
11 illness. Participants had blood drawn on 3 occasions: 30 days or less post-tick bite while
12 antibiotic-treatment-naïve, 6 to 8 weeks post tick bite, and 16 to 24 weeks post tick bite. Whole
13 blood samples were aliquoted, frozen, and stored at -80°C after collection. At each visit,
14 information was captured regarding symptoms, date of symptom onset, treatment status,
15 treatment regimen, and lab results. Specimens were received at Boca Biolistics Reference
16 Laboratory (Pompano Beach, FL, USA) and characterized for relevant tick-borne pathogens,
17 including *Bb*, *Babesia microti*, *Ehrlichia chaffeensis*, and *Anaplasma phagocytophilum*. Testing
18 was performed in-house and at ARUP Laboratories (Salt Lake City, UT, USA) on matched
19 serum collected from donors. DiaSorin and Immunonetics Lyme antibody testing was performed
20 at Boca Biolistics Reference Laboratory, and IgM- and IgG-specific antibody screening
21 for *Bb*, *B. microti*, *E. chaffeensis*, and *A. phagocytophilum* was performed at ARUP Laboratories.
22 For the present study, immunosequencing was performed on the first available sample from 18

1 donors who were seropositive for *Bb* by ELISA and immunoblot (either IgG or IgM), all of
2 whom were classified as STTT-positive by 2-tiered testing criteria.

3

4 *JHU cohort*

5 The SLICE study is a longitudinal, prospective cohort study that enrolled patients 18 years of age
6 or older with early LD who were self-referred or recruited from primary or urgent care settings
7 from 2008 to 2020. Eligible participants were enrolled primarily at study sites in Maryland, with
8 a small number enrolled at a satellite site in Southeastern Pennsylvania. At enrollment,
9 participants were required to have a visible EM of at least 5 cm in diameter diagnosed by a
10 healthcare provider. All patients had received no more than 72 hours of appropriate antibiotic
11 treatment for early LD at enrollment. Additional details and baseline clinical characteristics have
12 been previously published (48). Participants without a clinical or serologic history of LD were
13 recruited from similar primary care settings or through the community using flyers and online
14 advertising to serve as controls from Lyme-endemic areas. Controls were required to be STTT-
15 negative at the time of enrollment and at all subsequent visits. Participants in both groups were
16 excluded for a range of self-reported prior medical conditions paralleling those listed in the
17 proposed case definition for post-treatment LD syndrome (49), specifically, chronic fatigue
18 syndrome, fibromyalgia, unexplained chronic pain, sleep apnea or narcolepsy, autoimmune
19 disease, chronic neurologic disease, liver disease, hepatitis, HIV, cancer or malignancy in the
20 past 2 years, major psychiatric illness, or drug or alcohol abuse.

21 Patients with early LD were treated with 3 weeks of oral doxycycline (5) and seen regularly over
22 the course of 1 to 2 years. Samples collected before treatment, immediately after treatment, and 6
23 months post treatment were used for the present study, in addition to samples collected from

1 control participants at an initial study visit. Disseminated EM rash was defined as more than 1
2 rash site visible on physical exam at the pretreatment study visit, and local rash was defined as a
3 single EM rash site. Duration of illness was determined by self-report of the number of days
4 since the first appearance of LD-specific signs or symptoms. The number of LD symptoms at
5 enrollment was obtained through an interviewer-administered questionnaire. STTT status was
6 determined by Quest Diagnostics at each study time point using CDC recommendations
7 incorporating duration of illness at time of testing (7).

8 High-resolution HLA class I/II typing for the JHU cohort (cases only) was performed by Scisco
9 Genetics, Inc., (Seattle, WA, USA) using the ScisGo HLA v6 typing kit, as previously described
10 (50, 51).

11

12 *Database controls*

13 A total of 7,959 repertoires sampled during previous studies were selected from our database.
14 Inclusion was determined at the cohort level and based on the size of the cohort, geographic
15 region (United States and LD-endemic regions of Europe), and sequencing date (2019 or later, to
16 ensure consistent lab sequencing protocols). Samples were classified as being either from
17 endemic regions (Germany, Italy, or upper Midwest or Northeast regions of the United States) or
18 non-endemic regions (other regions of the United States). All individuals in these cohorts were
19 not tested but were presumed to be LD-negative.

20

1 **Assignment of cohorts**

2 Training cases were drawn from the LDB and Boca cohorts (Tables 1 and 2). To enrich for cases
3 with a likely immune response and maximize our ability to detect LD-associated ESs, the
4 training set was limited to 72 STTT-positive cases (54 LDB, 18 Boca). Training controls
5 included 2,981 repertoires from individuals from non-endemic regions of the United States and
6 Europe previously collected as part of other studies and presumed to be LD-negative.

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

12 TCR assay sensitivity was evaluated in the JHU cohort. Repertoires sampled from 211
13 participants at the time of enrollment passed QC thresholds established after model training
14 described below. A subset of patients in the JHU cohort (n=161) had sequenced repertoires that
15 passed QC from samples collected before and after treatment and 6 months post treatment.

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

1 **TCR repertoire immunosequencing**

2 Immunosequencing of TCR β CDR3 was performed using the immunoSEQ[®] Assay (Adaptive
3 Biotechnologies, Seattle, WA, USA) and analyzed as previously described (52–54). For
4 additional details, see the Supplementary Materials and Methods.

6 **Identification of LD-associated ESs**

7 Public TCR β amino acid sequences associated with early LD were identified as described
8 previously (28). Briefly, one-tailed Fisher’s exact tests (FETs) were performed on all unique
9 TCR sequences to compare frequencies of ESs in early Lyme samples with those in presumed-
10 negative controls. Unique sequences were defined based on the V-gene, J-gene, and CDR3
11 amino acid sequence. The P -value threshold for including a TCR in the ES list was treated as a
12 hyperparameter and was selected to maximize model performance as described below. The
13 resulting set of FET-defined ESs for cohort \mathcal{C} are denoted $S_{fet}^{\mathcal{C}}$.

14 A recent study in CMV demonstrated that many TCRs that are not identified as significant ESs
15 in small training datasets based solely on FET may be selected in larger training sets (54). This
16 observation motivates a simple classification problem: prediction of whether a TCR will be
17 identified as an ES by FET when the dataset grows to a specified size. To this end, a logistic
18 regression model was fitted, where the training data were the set of TCRs and corresponding
19 features observed in a previously reported cohort \mathcal{C}_{CMV} labeled for CMV serostatus (28). The
20 dependent binary variable was defined as 1 if the TCR was observed as in $S_{fet}^{\mathcal{C}^*}$ for some large
21 CMV-labeled cohort \mathcal{C}_{CMV}^* , and 0 otherwise. For each TCR, the following features were defined:
22 (1) average and maximum convergent recombination (CR) for cases and controls; (2) average
23 and maximum productive frequency for cases and controls; and (3) the number of sequences in

1 S_{fet} similar to the TCR, defined as sharing the same V-gene, having identical CDR3 length, and
2 differing by 1 amino acid.

3 In practice, a larger CMV-labeled cohort was unavailable. However, as more than 50% of North
4 American and European populations are seropositive for CMV (55), we applied a pseudo-
5 labeling procedure to construct C_{CMV}^* . Briefly, a logistic regression classifier (as defined in [44])
6 based on $S_{fet}^{C_{CMV}}$ was trained on the labeled CMV cohort C_{CMV} , then applied to all samples in the
7 LD training cohort. The inferred CMV status was then treated as observed and combined with
8 C_{CMV} , resulting in C_{CMV}^* , which was used to define $S_{fet}^{C_{CMV}^*}$. The resulting logistic regression classifier
9 was able to accurately predict which TCRs observed in C_{CMV} but not in $S_{fet}^{C_{CMV}}$ would end up in
10 $S_{fet}^{C_{CMV}^*}$ (area under the receiver operating characteristic [ROC] curve=0.84 in cross-validation; data
11 not shown). Features receiving the greatest weight in this model were the CR counts in cases
12 (likely indicating substantial clonal expansion) and the number of similar sequences in
13 $S_{fet}^{C_{CMV}}$ (likely indicating that the TCR responds to the same antigen as another ES).

14 The model fitted to CMV was used to infer ESs for LD. Combining these inferred ESs with ESs
15 identified by FET resulted in the final set of LD-associated ESs, $S^{C_{Lyme}}$. Hyperparameters in this
16 model were chosen using cross-validation in the context of the disease classification model
17 described below.

18

19 **Inferring early LD status based on ES counts in TCR repertoires**

20 Given a set of ESs S , the pair (y_i, x_i) can then be defined for each repertoire i , where x_i is the total
21 number of unique productive DNA TCR rearrangements in the sampled repertoire, and $y_i < x_i$ is
22 the number of those rearrangements that encode any of the ESs in S . If y_i is treated as sampled

1 from a random variable Y , the expected value of Y given x can be considered. By the way ESs are
2 defined, the distribution of $Y|x$ is expected to vary substantially between cases and controls.
3 While this could be treated as a classification problem to maximize the separation between cases
4 and controls (as in [39, 40]), $Y|x$ was instead explicitly modeled among control samples, with
5 classification based on standard units of deviation above and below expectation. This approach
6 provides superior control of specificity across populations given the extremely unbalanced nature
7 of our case/control dataset. To model this distribution, Y was assumed to follow a binomial
8 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
9 maximum number of ESs observed in any training repertoire, and

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

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

$$13 \quad \text{ModelScore}(y_i, x_i) = \frac{y_i - f(x_i)}{\sigma(x_i)}.$$

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

16 The observed data are moderately overdispersed with respect to the estimated variance (Figure
17 1A). As such, the final call threshold t was chosen to fix the prespecified false-positive rate of
18 1% on a set of 2,627 presumed LD-negative control samples as described above.

19 For a detailed description of TCR repertoire QC criteria, see the Supplementary Materials and
20 Methods.

21

1 **Antigen mapping and HLA assignment**

2 The MIRA assay was used for antigen mapping as described previously (32). To assign HLA
3 subtypes, ESs were clustered by a simple TCR amino acid similarity metric. ESs were
4 considered to belong to a cluster if they shared a V-gene family and if all members of a cluster
5 were connected by no more than a 1-Hamming difference between CDR3 regions. HLA subtypes
6 were inferred based on the results of one-tailed FETs performed between each ES and every
7 HLA subtype. For a detailed description of MIRA experiments and assignment of ESs to
8 antigens and HLA subtypes, see the Supplementary Materials and Methods.

9

10 **Statistical analysis**

11 Detailed statistical analyses associated with development of the TCR classifier and MIRA-based
12 TCR–antigen assignment are described above. Additional statistical analyses were performed
13 using the Python packages, *scipy* (version 1.5.4) and *statsmodels* (version 0.12.2). Significant
14 associations between TCR score and clinical variables were assessed by the Mann-Whitney *U*
15 test in univariate analysis and by multivariable linear regression with sex, age, and serostatus as
16 variables. *P* values <0.05 were considered significant. The correlation between TCR score and
17 the number of LD-related symptoms was assessed by Spearman’s rank-order correlation. For
18 comparisons of sensitivity, error bars represent mean \pm 95% CI by bootstrap sampling. For box-
19 and-whisker plots, boxes indicate median \pm interquartile ranges (IQR), and whiskers denote 1.5
20 times the IQR above the high quartile and below the low quartile.

21

1 **Data availability**

2 Deidentified data are available by request from AdaptiveBiotechnologies Medical Information
3 (<https://www.adaptivebiotech.com/medical-information-request/>).

4
5 **Acknowledgments**

6 We are grateful to the research participants who contributed samples and data used in this study
7 and to the physicians/health care providers who facilitated recruitment.

8 Funding for this study was provided by Adaptive Biotechnologies. Institutional support was
9 provided by the Steven and Alexandra Cohen Foundation (A.W.R., M.J.S., E.J.H., J.N.A.),
10 Global Lyme Alliance (A.W.R., J.N.A.), National Institutes of Health grant P30 AR070254
11 (M.J.S.), and Bay Area Lyme Foundation (E.J.H.).

12 Medical writing and editorial support were provided by Melanie Styers and Rachel Salmon of
13 BluPrint Oncology Concepts and Kristin MacIntosh and Shahin Shafiani of Adaptive
14 Biotechnologies.

15

1 **SUPPLEMENTARY MATERIALS**

2 Materials and Methods

3 Figure S1. Immunosequencing input DNA distributions by cohort.

4 Figure S2. Clinical correlates of TCR scoring in STTT+ and STTT- individuals.

5 Figure S3. Receiver operating characteristic curves using all control samples from Lyme-
6 endemic regions from Figure 2A as negatives.

7 Table S1. Multiple logistic regression of TCR model score on clinical features.

8 Table S2. Counts of enhanced sequences mapped to each protein by MIRA.

9

10

1 **REFERENCES**

- 2 1. Skar GL, Simonsen KA. 2022. Lyme disease, p1. *In* StatPearls [Internet]. StatPearls
3 Publishing, Treasure Island, FL.
- 4
- 5 2. Schwartz AM, Kugeler KJ, Nelson CA, Marx GE, Hinckley AF. 2021. Use of commercial
6 claims data for evaluating trends in Lyme disease diagnoses, United States, 2010–2018. *Emerg*
7 *Infect Dis* 27:499–507.
- 8
- 9 3. Kugeler KJ, Schwartz AM, Delorey MJ, Mead PS, Hinckley AF. 2021. Estimating the
10 frequency of Lyme disease diagnoses, United States, 2010–2018. *Emerg Infect Dis* 27:616–619.
- 11
- 12 4. Steere AC, Strle F, Wormser GP, Hu LT, Branda JA, Hovius JWR, Li X, Mead PS. 2016.
13 Lyme borreliosis. *Nat Rev Dis Prim* 2:16090.
- 14
- 15 5. Lantos PM, Rumbaugh J, Bockenstedt LK, Falck-Ytter YT, Agüero-Rosenfeld ME,
16 Auwaerter PG, Baldwin K, Bannuru RR, Belani KK, Bowie WR, Branda JA, Clifford DB,
17 DiMario FJ, Halperin JJ, Krause PJ, Lavergne V, Liang MH, Cody Meissner H, Nigrovic LE,
18 Nocton J (Jay) J, Osani MC, Pruitt AA, Rips J, Rosenfeld LE, Savoy ML, Sood SK, Steere AC,
19 Strle F, Sundel R, Tsao J, Vaysbrot EE, Wormser GP, Zemel LS. 2021. Clinical practice
20 guidelines by the Infectious Diseases Society of America (IDSA), American Academy of
21 Neurology (AAN), and American College of Rheumatology (ACR): 2020 guidelines for the
22 prevention, diagnosis, and treatment of Lyme disease. *Arthritis Care Res* 73:1–9.

23

- 1 6. Wormser GP, Dattwyler RJ, Shapiro ED, Halperin JJ, Steere AC, Klempner MS, Krause PJ,
2 Bakken JS, Strle F, Stanek G, Bockenstedt L, Fish D, Dumler JS, Nadelman RB. 2006. The
3 clinical assessments treatment, and prevention of Lyme disease, human granulocytic
4 anaplasmosis, and babesiosis: clinical practice guidelines by the Infectious Diseases Society of
5 America. *Clin Infect Dis* 43:1089–1134.
6
- 7 7. Centers for Disease Control and Prevention. 1995. Recommendations for test performance and
8 interpretation from the Second National Conference on Serologic Diagnosis of Lyme Disease.
9 *MMWR Morb Mortal Wkly Rep* 44:590–591.
10
- 11 8. Sigal LH. 1995. Diagnosis of Lyme disease. *JAMA* 274:1427–1428.
12
- 13 9. Fix AD, Strickland GT, Grant J. 1998. Tick bites and Lyme disease in an endemic: Setting
14 problematic use of serologic testing and prophylactic antibiotic therapy. *JAMA* 279:206–210.
15
- 16 10. Marques AR. 2018. Revisiting the Lyme disease serodiagnostic algorithm: the momentum
17 gathers. *J Clin Microbiol* 56:e00749-18.
18
- 19 11. Waddell LA, Greig J, Mascarenhas M, Harding S, Lindsay R, Ogden N. 2016. The accuracy
20 of diagnostic tests for Lyme disease in humans, a systematic review and meta-analysis of North
21 American research. *PLoS One* 11:e0168613.
22

- 1 12. Zweitzig D, Kopnitsky M, Zues Scientific. 2019. Validation of a modified two-tiered testing
2 (MTTT) algorithm for the improved diagnosis of Lyme disease [White paper]. Zeus Scientific,
3 Branchburg, NJ.
4
- 5 13. Branda JA, Strle K, Nigrovic LE, Lantos PM, Lepore TJ, Damle NS, Ferraro MJ, Steere AC.
6 2017. Evaluation of modified 2-tiered serodiagnostic testing algorithms for early Lyme disease.
7 *Clin Infect Dis* 64:1074–1080.
8
- 9 14. Nadelman RB, Nowakowski J, Forseter G, Goldberg NS, Bittker S, Cooper D, Agüero-
10 Rosenfeld M, Wormser GP. 1996. The clinical spectrum of early Lyme borreliosis in patients
11 with culture-confirmed erythema migrans. *Am J Med* 100:502–508.
12
- 13 15. Nowakowski J, Schwartz I, Liveris D, Wang G, Agüero-Rosenfeld MEA, Girao G, McKenna
14 D, Nadelman RB, Cavaliere LF, Wormser GP. 2001. Laboratory diagnostic techniques for
15 patients with early Lyme disease associated with erythema migrans: a comparison of different
16 techniques. *Clin Infect Dis* 33:2023–2027.
17
- 18 16. Schriefer ME. 2015. Lyme disease diagnosis: serology. *Clin Lab Med* 35:797–814.
19
- 20 17. Steere AC, McHugh G, Damle N, Sikand VK. 2008. Prospective study of serologic tests for
21 Lyme disease. *Clin Infect Dis* 47:188–195.
22

- 1 18. John TM, Taege AJ. 2019. Appropriate laboratory testing in Lyme disease. *Cleve Clin J Med*
2 86:751–759.
- 3
- 4 19. Seriburi V, Ndukwe N, Chang Z, Cox ME, Wormser GP. 2012. High frequency of false
5 positive IgM immunoblots for *Borrelia burgdorferi* in clinical practice. *Clin Microbiol Infect*
6 18:1236–1240.
- 7
- 8 20. Dressier F, Yoshinari NH, Steere AC. 1991. The T-cell proliferative assay in the diagnosis of
9 Lyme disease. *Ann Intern Med* 115:533–539.
- 10
- 11 21. Vaz A, Glickstein L, Field JA, McHugh G, Sikand VK, Damle N, Steere AC. 2001. Cellular
12 and humoral immune responses to *Borrelia burgdorferi* antigens in patients with culture-positive
13 early Lyme disease. *Infect Immun* 69:7437–7444.
- 14
- 15 22. Bockenstedt LK, Wooten RM, Baumgarth N. 2021. Immune response to *Borrelia*: lessons
16 from Lyme disease spirochetes. *Curr Issues Mol Biol* 42:145–190.
- 17
- 18 23. Soloski MJ, Crowder LA, Lahey LJ, Wagner CA, Robinson WH, Aucott JN. 2014. Serum
19 inflammatory mediators as markers of human Lyme disease activity. *PLoS One* 9:e93243.
- 20
- 21 24. Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. 1999. A direct
22 estimate of the human $\alpha\beta$ T cell receptor diversity. *Science* 286:958–961.

- 1 25. Neller MA, Burrows JM, Rist MJ, Miles JJ, Burrows SR. 2013. High frequency of
2 herpesvirus-specific clonotypes in the human T cell repertoire can remain stable over decades
3 with minimal turnover. *J Virol* 87:697–700.
4
- 5 26. Robins HS, Campregher P V, Srivastava SK, Wachter A, Turtle CJ, Kahsai O, Riddell SR,
6 Warren EH, Carlson CS. 2009. Comprehensive assessment of T-cell receptor β -chain diversity in
7 $\alpha\beta$ T cells. *Blood* 114:4099–4107.
8
- 9 27. Venturi V, Price DA, Douek DC, Davenport MP. 2008. The molecular basis for public T-cell
10 responses? *Nat Rev Immunol* 8:231–238.
11
- 12 28. Emerson RO, DeWitt WS, Vignali M, Gravley J, Hu JK, Osborne EJ, Desmarais C, Klinger
13 M, Carlson CS, Hansen JA, Rieder M, Robins HS. 2017. Immunosequencing identifies
14 signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell
15 repertoire. *Nat Genet* 49:659–665.
16
- 17 29. Snyder MT, Gittelman MR, Klinger M, May HD, Osborne JE, Taniguchi R, Jabran Zahid H,
18 Kaplan MI, Dines NJ, Noakes TM, Pandya R, Chen X, Elasady S, Svejnoha E, Ebert P, Pesesky
19 WM, De Almeida P, O'Donnell H, DeGottardi Q, Keitany G, Lu J, Vong A, Elyanow R, Fields
20 P, Greissl J, Baldo L, Semprini S, Cerchione C, Nicolini F, Mazza M, Delmonte M Ottavia,
21 Dobbs K, Laguna-Goya R, Carreño-Tarragona G, Barrio S, Imberti L, Sottini A, Quiros-Roldan
22 E, Rossi C, Biondi A, Bettini RL, D'Angio M, Bonfanti P, Tompkins F. Miranda, Alba C,
23 Dalgard C, Sambri V, Martinelli G, Goldman DJ, Heath RJ, Su CH, Notarangelo DL, Paz-Artal

- 1 E, Martinez-Lopez J, Carlson M Jonathan, Robins SH. 2020. Magnitude and dynamics of the T-
2 cell response to SARS-CoV-2 infection at both individual and population levels. medRxiv
3 <https://doi.org/10.1101/2020.07.31.20165647>.
4
- 5 30. Elyanow R, Snyder TM, Dalai SC, Gittelman RM, Boonyaratanakornkit J, Wald A, Selke S,
6 Wener MH, Morishima C, Greninger AL, Gale Jr. M, Hsiang T-Y, Jing L, Holbrook MR,
7 Kaplan IM, Zahid HJ, May DH, Carlson JM, Baldo L, Manley T, Robins HS, Koelle DM. 2022.
8 T-cell receptor sequencing identifies prior SARS-CoV-2 infection and correlates with
9 neutralizing antibodies and disease severity. JCI Insight 7:e150070.
10
- 11 31. Dalai SC, Dines JN, Snyder TM, Gittelman RM, Eerkes T, Vaney P, Howard S, Akers K,
12 Skewis L, Monteforte A, Witte PR, Wolf C, Nesse H, Herndon M, Qadeer J, Duffy S, Svejnoha
13 E, Taromino C, Kaplan IM, Alsobrook J, Manley T, Baldo L. 6 May 2022. Clinical validation of
14 a novel T-cell receptor sequencing assay for identification of recent or prior SARS-CoV-2
15 infection. Clin Infect Dis <https://doi.org/10.1093/cid/ciac353>.
16
- 17 32. Klinger M, Pepin F, Wilkins J, Asbury T, Wittkop T, Zheng J, Moorhead M, Faham M.
18 2015. Multiplex identification of antigen-specific T cell receptors using a combination of
19 immune assays and immune receptor sequencing. PLoS One 10:e0141561.
20
- 21 33. Gittelman RM, Lavezzo E, Snyder TM, Zahid HJ, Carty CL, Elyanow R, Dalai SC, Kirsch I,
22 Baldo L, Manuto L, Franchin E, Del Vecchio C, Pacenti M, Boldrin C, Cattai M, Saluzzo F,
23 Padoan A, Plebani M, Simeoni F, Bordini J, Lorè NI, Lazarević D, Cirillo DM, Ghia P, Toppo S,

- 1 Carlson JM, Robins HS, Crisanti A, Tonon G. 2022. Longitudinal analysis of T-cell receptor
2 repertoires reveals shared patterns of antigen-specific response to SARS-CoV-2 infection. *JCI*
3 *Insight* 7:e151849.
- 4
- 5 34. Koszinowski UH, Reddehase MJ, Jonjic S. 1991. The role of CD4 and CD8 T cells in viral
6 infections. *Curr Opin Immunol* 3:471–475.
- 7
- 8 35. Shepherd FR, McLaren JE. 2020. T cell immunity to bacterial pathogens: mechanisms of
9 immune control and bacterial evasion. *Int J Mol Sci* 21:1–32.
- 10
- 11 36. Marshall JS, Warrington R, Watson W, Kim HL. 2018. An introduction to immunology and
12 immunopathology. *Allergy Asthma Clin Immunol* 14:49.
- 13
- 14 37. Hastey CJ, Elsner RA, Barthold SW, Baumgarth N. 2012. Delays and diversions mark the
15 development of B cell responses to *Borrelia burgdorferi* infection. *J Immunol* 188:5612.
- 16 38. Aucott JN, Rebman AW, Crowder LA, Kortte KB. 2013. Post-treatment Lyme disease
17 syndrome symptomatology and the impact on life functioning: Is there something here? *Qual*
18 *Life Res* 22:75–84.
- 19
- 20 38. Aucott JN, Rebman AW, Crowder LA, Kortte KB. 2013. Post-treatment Lyme disease
21 syndrome symptomatology and the impact on life functioning: is there something here? *Qual*
22 *Life Res* 22:75–84.
- 23

- 1 39. Rebman AW, Crowder LA, Kirkpatrick A, Aucott JN. 2015. Characteristics of
2 seroconversion and implications for diagnosis of post-treatment Lyme disease syndrome: acute
3 and convalescent serology among a prospective cohort of early Lyme disease patients. Clin
4 Rheumatol 34:585–589.
- 5
- 6 40. Aguero-Rosenfeld ME, Nowakowski J, Bittker S, Cooper D, Nadelman RB, Wormser GP.
7 1996. Evolution of the serologic response to *Borrelia burgdorferi* in treated patients with culture-
8 confirmed erythema migrans. J Clin Microbiol 34:1–9.
- 9
- 10 41. McKisic MD, Barthold SW. 2000. T-cell-independent responses to *Borrelia burgdorferi* are
11 critical for protective immunity and resolution of Lyme disease. Infect Immun 68:5190–5197.
- 12
- 13 42. Pegalajar-Jurado A, Schriefer ME, Welch RJ, Couturier MR, MacKenzie T, Clark RJ,
14 Ashton L V., Delorey MJ, Molins CR. 2018. Evaluation of modified two-tiered testing
15 algorithms for Lyme disease laboratory diagnosis using well-characterized serum samples. J Clin
16 Microbiol 56:e01943-17.
- 17
- 18 43. Mosel M, Rebman A, Carolan H, Montenegro T, Lovari R, Schutzer S, Ecker D, Yang T,
19 Ramadoss N, Robinson W, Soloski M, Eshoo M, Aucott J. 2020. Molecular microbiological and
20 immune characterization of a cohort of patients diagnosed with early Lyme disease. J Clin
21 Microbiol 59:e00615-20.

22

- 1 44. Aucott JN, Yang T, Yoon I, Powell D, Geller SA, Rebman AW. 2022. Risk of post-treatment
2 Lyme disease in patients with ideally-treated early Lyme disease: a prospective cohort study. *Int*
3 *J Infect Dis* 116:230–237.
- 4
- 5 45. Kalish RA, McHugh G, Granquist J, Shea B, Ruthazer R, Steere AC. 2001. Persistence of
6 immunoglobulin M or immunoglobulin G antibody responses to *Borrelia burgdorferi* 10–20
7 years after active Lyme disease. *Clin Infect Dis* 33:780–785.
- 8
- 9 46. Zajkowska JM. 2014. Antibody-based techniques for detection of Lyme disease: a
10 challenging issue. *Antib Technol J* 4:33–44.
- 11
- 12 47. Horn E, Dempsey G, Schotthoefer A, Prisco U, McArdle M, Gervasi S, Golightly M, De
13 Luca C, Evans M, Pritt B, Theel E, Iyer R, Liveris D, Wang G, Goldstein D, Schwartz I. 2020.
14 The Lyme disease biobank: Characterization of 550 patient and control samples from the East
15 Coast and Upper Midwest of the United States. *J Clin Microbiol* 58:e00032-20.
- 16
- 17 48. Rebman AW, Yang T, Mihm EA, Novak CB, Yoon I, Powell D, Geller SA, Aucott JN. 2021.
18 The presenting characteristics of erythema migrans vary by age, sex, duration, and body location.
19 *Infection* 49:685–692.
- 20
- 21 49. Aucott J, Crowder L, Kortte K. 2013. Development of a foundation for a case definition of
22 post-treatment Lyme disease syndrome. *Int J Infect Dis* 17:e443-9.
- 23

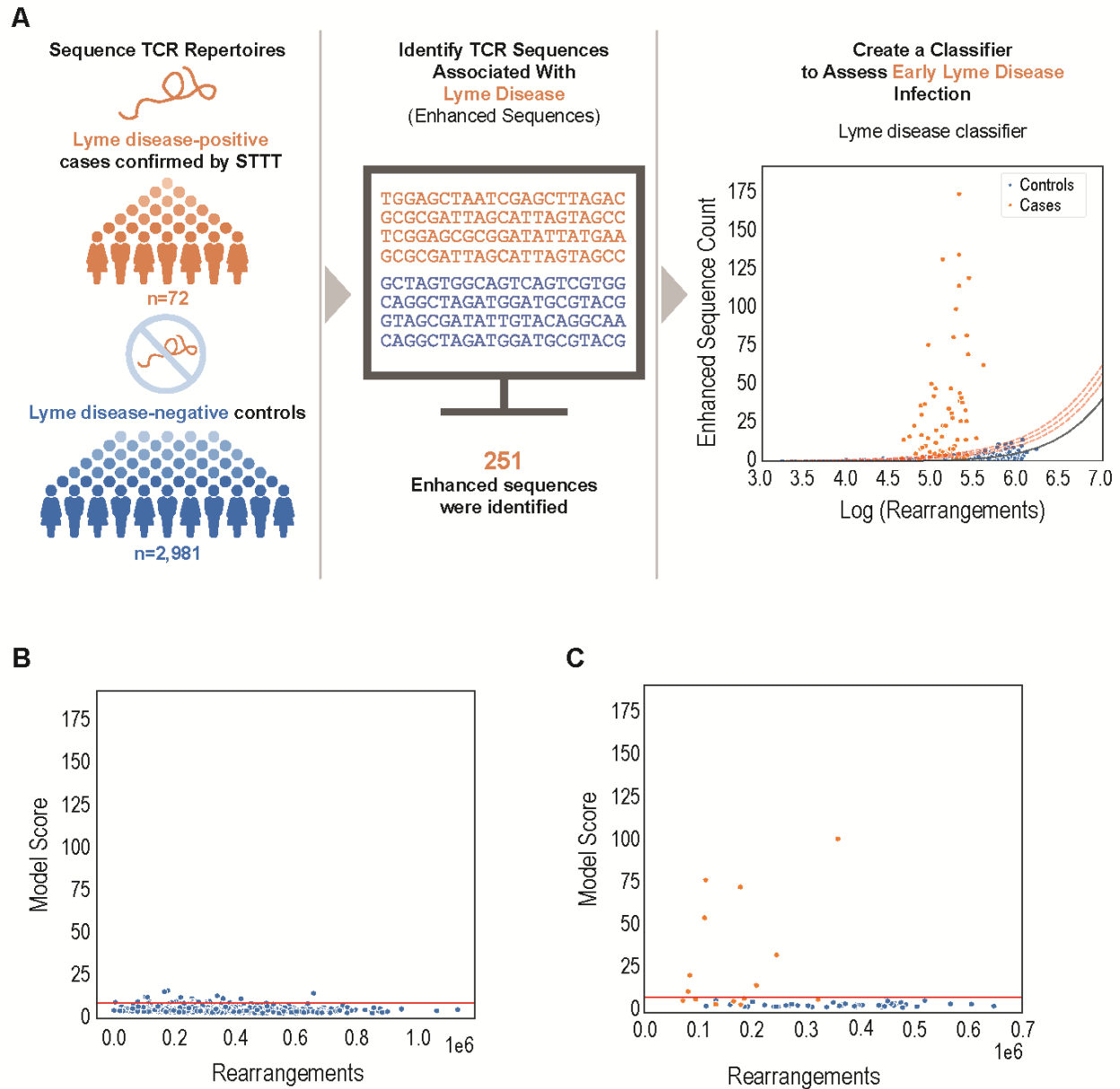
- 1 50. Nelson WC, Pyo CW, Vogan D, Wang R, Pyon YS, Hennessey C, Smith A, Pereira S,
2 Ishitani A, Geraghty DE. 2015. An integrated genotyping approach for HLA and other complex
3 genetic systems. *Hum Immunol* 76:928–938.
4
- 5 51. Smith AG, Pereira S, Jaramillo A, Stoll ST, Khan FM, Berka N, Mostafa AA, Pando MJ,
6 Usenko CY, Bettinotti MP, Pyo C-W, Nelson WC, Willis A, Askar M, Geraghty DE. 2019.
7 Comparison of sequence-specific oligonucleotide probe vs next generation sequencing for HLA-
8 A, B, C, DRB1, DRB3/B4/B5, DQA1, DQB1, DPA1, and DPB1 typing: toward single-pass
9 high-resolution HLA typing in support of solid organ and hematopoietic cell transplant
10 programs. *HLA* 94:296–306.
11
- 12 52. Robins H, Desmarais C, Matthis J, Livingston R, Andriesen J, Reijonen H, Carlson C,
13 Nepom G, Yee C, Cerosaletti K. 2012. Ultra-sensitive detection of rare T cell clones. *J Immunol*
14 *Methods* 375:14–19.
15
- 16 53. Carlson CS, Emerson RO, Sherwood AM, Desmarais C, Chung MW, Parsons JM, Steen MS,
17 LaMadrid-Herrmannsfeldt MA, Williamson DW, Livingston RJ, Wu D, Wood BL, Rieder MJ,
18 Robins H. 2013. Using synthetic templates to design an unbiased multiplex PCR assay. *Nat*
19 *Commun* 4:2680.
20
- 21 54. Pavlović M, Scheffer L, Motwani K, Kanduri C, Kompova R, Vazov N, Waagan K, Bernal
22 FLM, Costa AA, Corrie B, Akbar R, Al Hajj GS, Balaban G, Brusko TM, Chernigovskaya M,
23 Christley S, Cowell LG, Frank R, Grytten I, Gundersen S, Haff IH, Hovig E, Hsieh PH,

1 Klambauer G, Kuijjer ML, Lund-Andersen C, Martini A, Minotto T, Pensar J, Rand K, Riccardi
2 E, Robert PA, Rocha A, Slabodkin A, Snapkov I, Sollid LM, Titov D, Weber CR, Widrich M,
3 Yaari G, Greiff V, Sandve GK. 2021. The immuneML ecosystem for machine learning analysis
4 of adaptive immune receptor repertoires. *Nat Mach Intell* 2021 3:936–944.
5 55. Zuhair M, Smit GSA, Wallis G, Jabbar F, Smith C, Devleeschauwer B, Griffiths P. 2019.
6 Estimation of the worldwide seroprevalence of cytomegalovirus: A systematic review and meta-
7 analysis. *Rev Med Virol* 29:e2034.

8

9

1 FIGURES



2

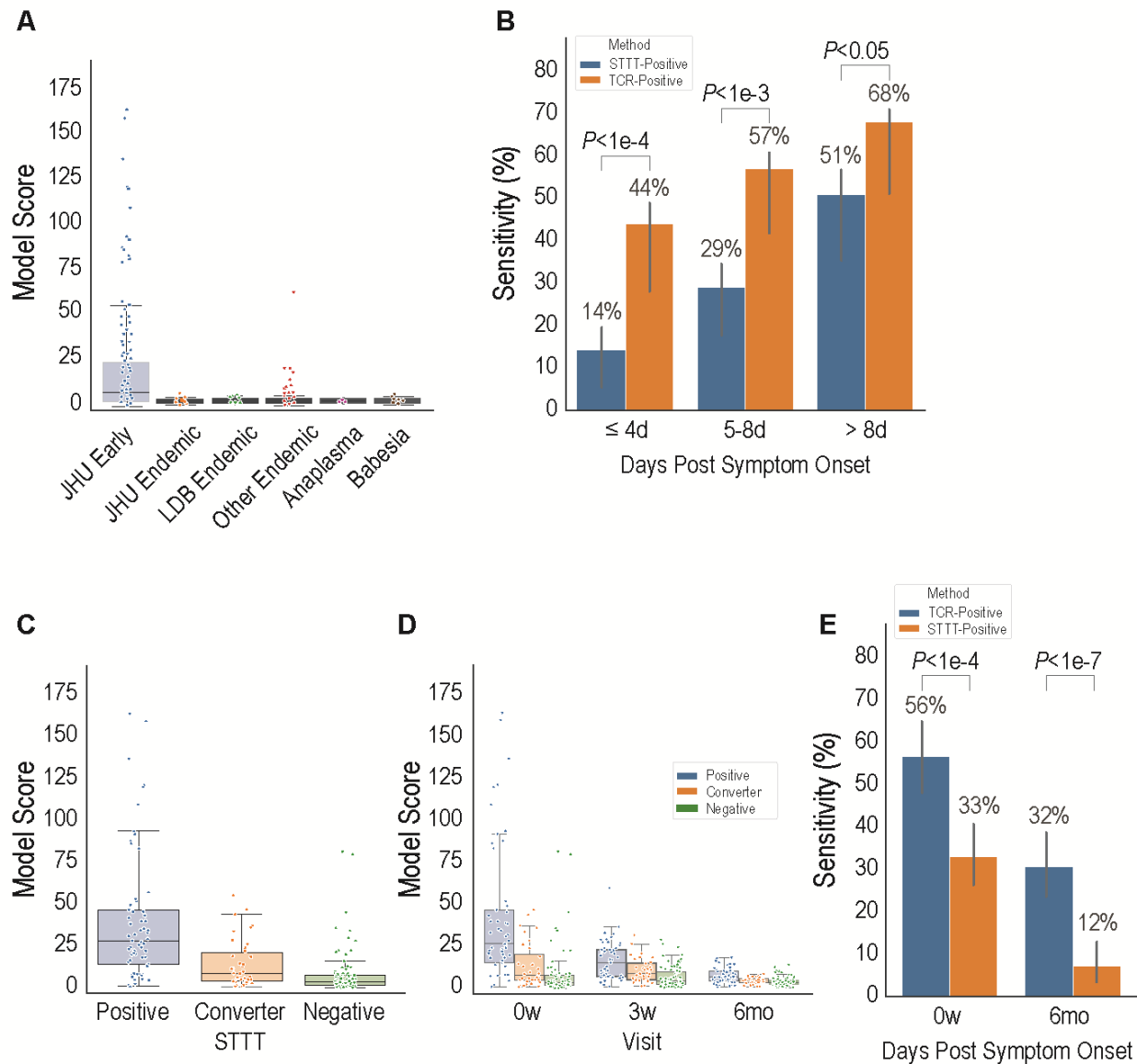
3 **Figure 1. LD-associated TCRs distinguish cases from controls in training cohorts. (A)**

4 Development and training of the classifier used to assess *Bb* infection using a T-cell receptor
5 assay. The classifier was trained using TCR repertoires sequenced from 72 STTT–confirmed
6 cases (orange) and 2,981 controls (blue) (see Table 2). A total of 251 LD-associated TCR
7 sequences (enhanced sequences) were identified. Distribution of the number of TCR

1 rearrangements encoding enhanced sequences as a function of the (log-transformed) total
2 number of unique TCR rearrangements identified in a repertoire. Distribution of enhanced
3 sequences in control samples approximately follows a logistic-growth curve (solid black line;
4 dashed red lines indicate +2, +3, and +4 standard deviations from fit), which was used to define a
5 scoring function. **(B)** Distribution of the model scores is largely invariant to the number of
6 unique rearrangements in an independent set of controls from Lyme-endemic areas (n=2,627).
7 Red line indicates 99th percentile distribution in this cohort (score=4.2675), which was defined
8 as the positive-call threshold. **(C)** Model score distribution in a holdout set of repertoires from
9 the LDB cohort (n=15 cases; n=48 controls), collected in 2019 and immunosequenced after
10 model training.

11

1



2

3 **Figure 2. Validation of the TCR classifier in the JHU cohort and other holdout endemic**

4 **controls. (A)** Model score distribution in early LD samples from JHU (blue, n=211), in addition

5 to holdout controls from Lyme-endemic areas. JHU (orange, n=45) and LDB (green, n=115)

6 controls from Lyme-endemic areas were LD-negative based on clinical assessment and negative

7 STTT testing. Other controls from Lyme-endemic areas (red, n=2,471) were drawn from our

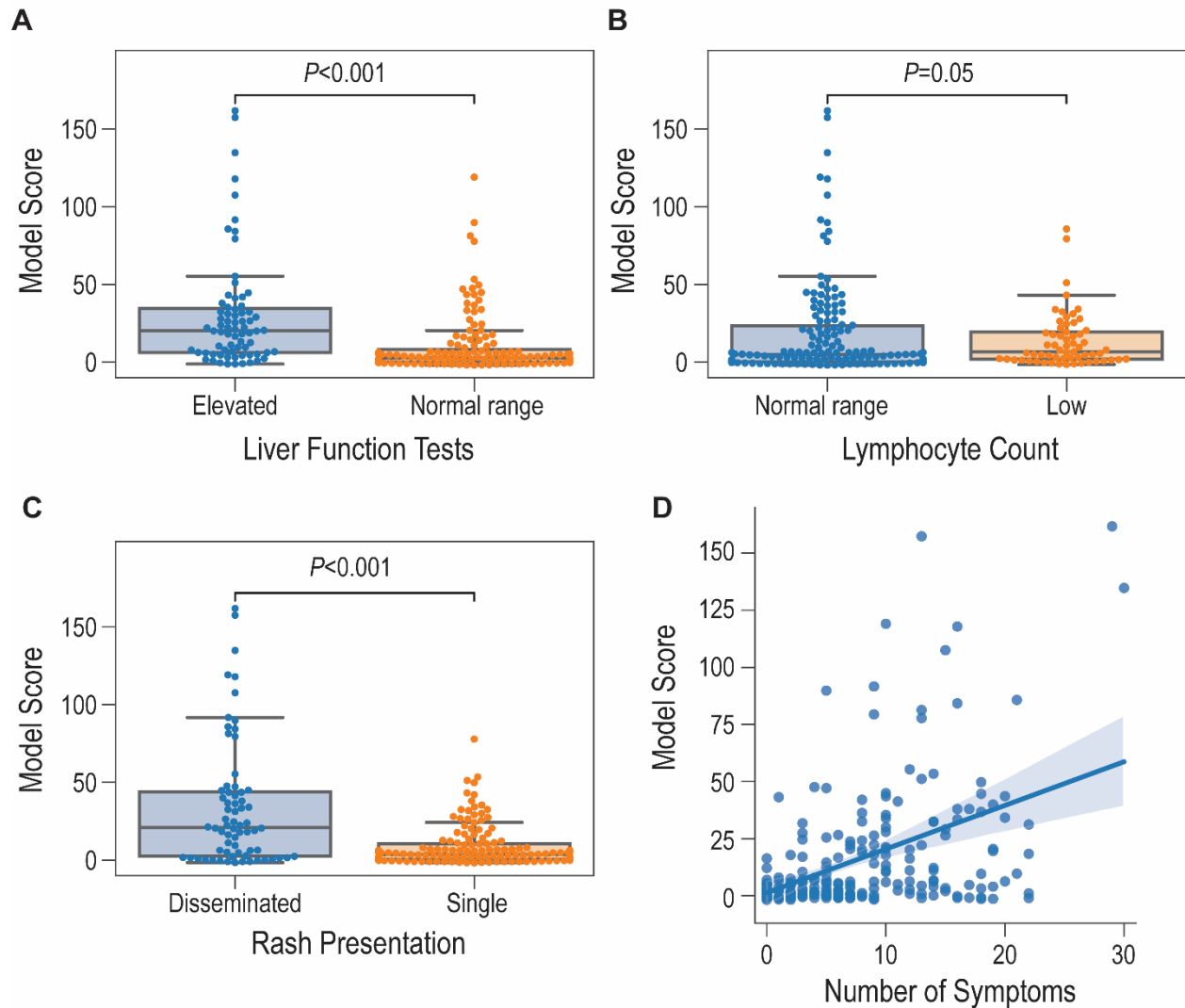
1 database of repertoires sampled from individuals in endemic regions in the United States and
2 Europe who were presumed negative for LD. The *Anaplasma* samples (purple, n=4) were from
3 the LDB cohort, and the *Babesia* samples were from the LDB and DLS cohorts (brown, n=5 and
4 n=12, respectively). **(B)** Sensitivity of STTT and the TCR classifier for individuals in the JHU
5 cohort, stratified by symptom duration (days) at time of enrollment, along with bootstrapped
6 95% CI. Participants were stratified based on self-reported symptom duration into bins of ≤ 4
7 days (n=73), 5-8 days (n=75) or >8 days (n=63). The sensitivities of both TCR testing and STTT
8 increased with longer duration of symptoms reported at the time of testing. **(C)** Model score
9 distribution for JHU early disease samples stratified by STTT serostatus at enrollment and
10 posttreatment follow-up. Positive (blue, n=64): STTT-positive at enrollment; converter STTT
11 (orange, n=38): STTT-negative at enrollment and STTT-positive at posttreatment follow-up;
12 negative (green, n=109): STTT-negative at both visits. **(D)** Longitudinal dynamics of TCR
13 scoring by serostatus for the JHU cohort. Positive (blue, n=53): STTT-positive at enrollment;
14 converter (orange, n=32): STTT-negative at enrollment and STTT positive at posttreatment
15 follow-up; negative (green, n=76): STTT-negative at both visits. **(E)** For the 163 patients who
16 had STTT performed pretreatment and at 6-month follow-up, difference in sensitivities was
17 calculated for recent Lyme infection between STTT and TCR test. In box-and-whisker plots,
18 boxes indicate median \pm IQR, and whiskers denote 1.5 times the IQR above the high quartile and
19 below the low quartile. Significant differences in sensitivity were evaluated by T test.

20

21

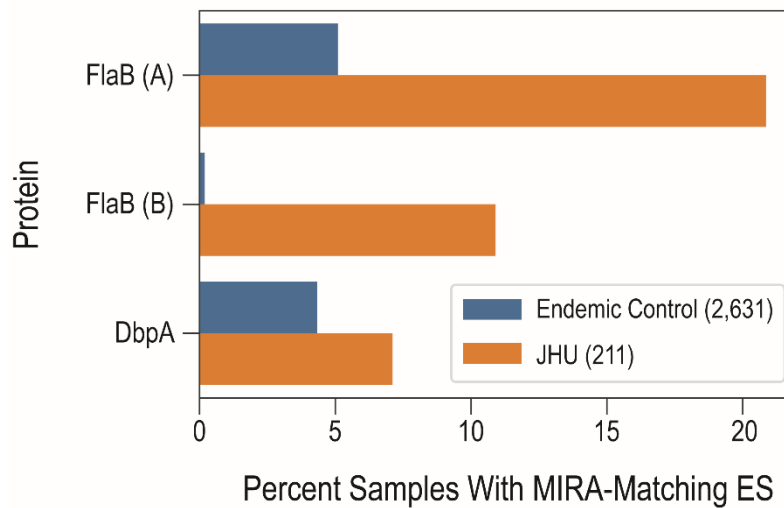
22

23



1
2 **Figure 3. Clinical correlates of TCR scoring.** TCR scores were stratified by (A) liver function
3 test results (elevated [n=72] vs normal [n=139]), (B) lymphocyte counts (normal [n=150] vs low
4 [n=61]), or (C) presentation of rash (multiple/disseminated [n=68] vs single [n=143]) and (D)
5 plotted as a function of the number of Lyme-related symptoms (Spearman $R^2=0.17$. P values,
6 Mann-Whitney U test). In box-and-whisker plots, boxes indicate median \pm IQR, and whiskers
7 denote 1.5 times the IQR above the high quartile and below the low quartile.

8



1
2 **Figure 4. Percentage of early LD samples with ESs assigned by MIRA to the indicated *Bb***
3 **antigens. FlaB (A/B), flagellin protein B antigen A/B; DbpA, decorin-binding protein A.**
4

1 **TABLES**

2 **Table 1. Description and criteria for cohorts used in study.**

3

Name of Cohort	Cohort	Description/Criteria
LDB (pre-2019)	Case	<ul style="list-style-type: none"> Patients presenting prior to 2019: <ul style="list-style-type: none"> With EM >5 cm in diameter or an erythematous, annular, expanding skin lesion ≤5 cm, diagnosed by a health care provider at study enrollment or With no history of chronic fatigue syndrome, rheumatologic disease, or multiple sclerosis 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 STTT-positive
	Control (endemic)	<ul style="list-style-type: none"> Healthy individuals enrolled prior to 2019 living in an area of endemicity with no history of Lyme disease or tick-borne infection STTT-negative
	Cross-reactivity case	<ul style="list-style-type: none"> Symptomatic patients STTT-negative PCR-positive for babesiosis (n=5) or anaplasmosis (n=4)
Boca Biologics	Case	<ul style="list-style-type: none"> Patients with an EM rash, positive serology results, and/or evidence of a tick bite STTT-positive
JHU	Case	<ul style="list-style-type: none"> Patients presenting with EM ≥5 cm diagnosed by a health care provider at study enrollment Had received ≤72 hours of appropriate antibiotic treatment for early Lyme disease at enrollment
	Control (endemic)	<ul style="list-style-type: none"> No clinical or serologic history of Lyme disease STTT-negative
LDB (2019)	Case	<ul style="list-style-type: none"> Patients presenting during 2019 with the same characteristics described for the LDB case subgroup, except positives were confirmed by STTT and/or PCR
	Control (endemic)	<ul style="list-style-type: none"> Healthy individuals enrolled during 2019 with the same characteristics described for the LDB controls from Lyme-endemic areas subgroup
Database	Control (endemic)	<ul style="list-style-type: none"> Individuals with unknown Lyme disease status living in Lyme-endemic regions in the United States and Europe
Database	Control (non-endemic)	<ul style="list-style-type: none"> Individuals with unknown Lyme disease status living in Lyme non-endemic regions in the United States
DLS cohort	Cross-reactive	<ul style="list-style-type: none"> Suspected <i>Babesia</i> samples (PCR-positive)

4
5
6
7
8

1 **Table 2. Cohorts used for training, setting the classification threshold, verification and**
 2 **determination of sensitivity and specificity of the classifier.**

3

Study Phase	Name of Cohort	Lyme Disease Status	Samples Used
Classifier training	LBD (pre-2019)	Case	54
	Boca Biolistics	Case	18
	In-house database	Control (non-endemic)	2,981
Setting threshold	In-house database	Control (endemic)	2,507
	LDB (pre-2019)	Control (endemic)	120
Verification (holdout set)	LDB (2019)	Case	15
	LDB (2019)	Control (endemic)	48
Sensitivity	JHU	Case	211
Specificity	LDB (pre-2019)	Control (endemic)	115
	In-house database	Control (endemic)	2,471
	JHU	Control (endemic)	45
	DLS	Control (suspected anaplasmosis)	12
Cross-reactivity	LDB (pre-2019)	Cross-reactivity (endemic) STTT-negative, PCR-confirmed cases of anaplasmosis and babesiosis	9

4

5

6

7

8

9

1 **Table 3. TCR classifier sensitivity stratified by serostatus in the JHU cohort.**

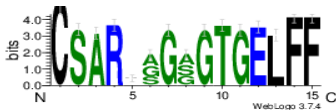

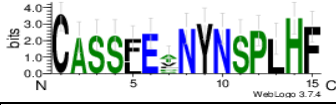

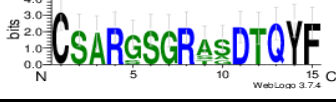
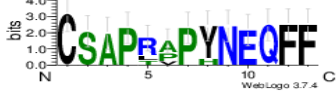
2

Serostatus	Sensitivity (95% CI ^a)	TCR-Positive (n/N)
All	0.56 (0.49–0.62)	118/211
STTT-positive	0.92 (0.84–0.98)	59/64
STTT posttreatment seroconverter	0.58 (0.45–0.74)	22/38
Persistent STTT-negative	0.34 (0.27–0.43)	37/109

3 ^aBy bootstrap sampling.

4

1 **Table 4. TCR clusters as defined by connected components within a 1–amino-acid**
 2 **change in CDR3 and the same V-gene family.**
 3

Cluster	CDR3 Motif	V Family	CDR3 Length (bp)	No. of Sequences	HLA	MIRA-Match TCRs
1		V20	15	65	DPA1*01:03+ DPB1*04:01	0
2		V06	14	12	ND	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])

4 bp, base pair

5

6

7

8

9

10

11

12

13

1 **SUPPLEMENTARY MATERIALS**

2

3 **MATERIALS AND METHODS**

4 **TCR repertoire immunosequencing**

5 Immunosequencing of TCR β CDR3 was performed using the immunoSEQ[®] Assay (Adaptive
6 Biotechnologies, Seattle, WA, USA). Extracted genomic DNA was amplified in a bias-
7 controlled multiplex PCR, followed by high-throughput sequencing. Sequences were collapsed
8 and filtered to identify and quantitate the absolute abundance of each unique TCR β CDR3 region
9 for further analysis as previously described (26, 52, 53). Sequencing reactions contained a
10 median of 7,884.0 ng of input DNA (range, 239.9–55,186.4 ng) and yielded a median of 314,948
11 T-cell templates per sample (range, 15–1,837,496) (Figure S1). The T-cell fraction (percentage
12 of T cells among the estimated number of nucleated cells input) ranged from 0.3% to 90%
13 (median, 26.0%).

14

15 **TCR repertoire QC criteria**

16 The 2 key parameters of the classifier are the number of unique productive rearrangements, x ,
17 and the number of unique productive rearrangements encoding an ES, y . For a given blood
18 sample, the value of x is determined by the quantity of DNA, the fraction of cells that are T cells,
19 and the diversity of T cells. In rare cases, x is too small to yield meaningful information, or
20 significantly larger than observed in our training data, making extrapolation of $Y|x$ problematic.
21 Therefore, acceptance criteria were predefined for the number of unique productive
22 rearrangements based on the observed distribution of x in the training data.

1 The information contained in ESs is asymmetric: for small x , large y is considered to be evidence
2 of LD, while small y may simply reflect a lack of sequenced T cells. Thus, QC criteria were
3 treated asymmetrically. Specifically, x_{\max} and x_{\min} were defined as the upper and lower QC
4 thresholds, which were prespecified to be equal to the 1st and 99th percentiles, respectively, of x
5 observed in the training data. A sample i then failed QC if $x_i > x_{\max}$, or if $x_i < x_{\min}$ and
6 $ModelScore(y_i, x_i) < t$.

7

8 **Determination of antigen specificity of Lyme-associated TCRs**

9 To evaluate the potential specificity of antigens recognized by TCRs in the LD classifier, ESs
10 were clustered by sequence similarity. Statistical assignment of individual TCRs to HLA
11 subtypes resulted in a consistent HLA assignment for the cluster. To further characterize TCR–
12 antigen specificity, MIRA was used to identify TCR epitopes, as detailed in the next section.
13 Query peptides (777 total) were derived from 26 *Bb* proteins, and individual peptides or groups
14 of related peptides were assigned to 1 of 426 unique MIRA pools. MIRA was then performed on
15 T cells derived from peripheral blood mononuclear cells collected from 395 healthy individuals
16 using a version of the assay that selects for HLA-II–restricted CD4+ T cells. Basic Local
17 Alignment Search Tool (BLAST) analysis of the target epitope was used to compare *Bb* FlaB to
18 other *Borrelia* species and other pathogens. Specificity of our approach was further evaluated by
19 comparing the Lyme-associated ESs with a set of TCRs from 507 individuals that were
20 previously mapped to 325 SARS-CoV-2 antigen pools by MIRA (29 and Supplemental reference
21 1).

22

1 **Antigen-stimulation experiments (MIRA assay)**

2 *Panel design*

3 The multiplex identification of antigen-specific TCRs (MIRA) assay was set up, performed, and
4 analyzed as described previously (32). Briefly, 2 panels of peptides were designed and tested in
5 the assay. The Lyme-MIRA1 panel used antigens that are known to be presented in LD,
6 including tiled portions of the DbpA, OspC, OspA, BBK32, BBA52, and VlsE proteins
7 (Supplemental reference 2). The Lyme-MIRA2 panel used peptides derived from antigens
8 presented via HLA class II upon *Bb* infection. The *Bb*-derived antigens include elongation factor
9 Tu (WP_002657015.1), BB_0418 (WP_002658797.1), p83/100 (CAA57125.1), ABC transporter
10 (PRR58667.1), lipoprotein LA7 (WP_002657819.1), GAPDH (AAB53930.1), chaperonin
11 GroEL (WP_002657108.1), flagellin (WP_002661938.1), OspA (WP_010890378.1), and p66
12 (WP_002656762.1). The peptide tiling strategy was used across the entirety of each antigen,
13 yielding a series of peptides, each 17 amino acids (aa) long with a 7-aa overlap between peptides.
14 The peptides were pooled in a combinatorial fashion as described previously (32); peptides that
15 were overlapping or in close proximity in the viral proteome were grouped together into antigen
16 sets. Each antigen set was then placed in a subset of 5 unique pools out of 11 total pools in the
17 Lyme-MIRA1 panel, or 6 pools out of 12 in the Lyme-MIRA2 panel, referred to as its
18 occupancy.

19

20 *Naïve antigen-stimulation experiments*

21 A total of 304 experiments were run with the Lyme-MIRA1 panel (all “naïve” experiments, see
22 below) and 174 with the Lyme-MIRA2 panel. For the “naïve” experiments, CD14+ monocytes

1 were selected from peripheral blood mononuclear cells (PBMCs; Miltenyi Biotech, Auburn, CA,
2 USA) and stimulated with granulocyte-macrophage colony-stimulating factor (GM-CSF) and
3 interleukin (IL)-4 (BioLegend, San Diego, CA, USA) to drive dendritic cell (DC) differentiation
4 *in vitro*. On day 3, GM-CSF, IL-4, interferon- γ (IFN- γ ; BioLegend, San Diego, CA, USA), and
5 lipopolysaccharide (LPS; Sigma-Aldrich, St. Louis, MO, USA or eBioscience, Inc, San Diego,
6 CA, USA) were added to promote DC maturation. Also on day 3, naïve T cells were isolated
7 from PBMCs (StemCell, Vancouver, BC, Canada) and incubated overnight with IL-7. On day 4,
8 naïve T cells were combined with the differentiated CD14⁺ monocytes, IL-21 (BioLegend, San
9 Diego, CA, USA), and a pool of all peptides present in the panel to be used for restimulation.
10 Cultures were supplemented with IL-7, IL-15, and IL-2 every 2 to 3 days for an additional 12 to
11 14 days. Cells harvested from the expansion culture were divided into a series of replicate
12 cultures, and each was restimulated using a distinct peptide pool from the panel under
13 investigation. After incubation at 37°C for ~20 hours, each culture was stained with antibodies
14 (BioLegend, San Diego, CA, USA) for sorting by flow cytometry. Cells were then washed and
15 suspended in phosphate-buffered saline containing 2% fetal bovine serum (FBS), 1 mM EDTA,
16 and 4,6-diamidino-2-phenylindole (DAPI) for exclusion of non-viable cells. Cells were acquired
17 and sorted using a FACSMelody (BD Biosciences, Franklin Lakes, NJ, USA) instrument.
18 Sorted antigen-specific (CD4⁺CD137⁺CD145⁺, CD25^{lo}) T cells were pelleted and lysed in RLT
19 Plus buffer (Qiagen, Germantown, MD, USA) for nucleic acid isolation.

20

21 *Assignment of ESs to antigens*

22 To assign ESs to antigens, RNA was isolated using AllPrep DNA/RNA mini and/or micro kits,
23 according to the manufacturer's instructions (Qiagen). RNA was then reverse transcribed to

1 cDNA using Vilo kits (Life Technologies, Carlsbad, CA, USA), and TCR β amplification was
2 performed using the immunoSEQ Assay described above.

3 After immunosequencing, the behavior of T-cell clonotypes was examined by tracking read
4 counts across each sorted pool. True antigen-specific clones should be specifically enriched in a
5 unique occupancy pattern corresponding to the presence of 1 of the query antigens in 5 or 6
6 pools in the Lyme-MIRA1 and Lyme-MIRA2 panels. Methods used to assign antigen specificity
7 to TCR clonotypes have been reported previously (32). In addition to these methods, a non-
8 parametric Bayesian model was developed to compute the posterior probability that a given
9 clonotype was antigen-specific. This model uses the available read counts of TCRs to estimate a
10 mean-variance relationship within a given experiment, as well as the probability that a clone will
11 have zero read counts due to incomplete sampling of low-frequency clones. Together, this model
12 considers the observed read counts of a clonotype across all pools and estimates the posterior
13 probability of a clone responding to all valid addresses and an additional hypothesis that a clone
14 is activated in all pools (truly activated, but not specific to any of our query antigens). To define
15 antigen-specific clones, we identified TCR clonotypes assigned to a query antigen from this
16 model with a posterior probability ≥ 0.7 .

17 TCR sequences from MIRA were compared to the ES list based on V-gene, J-gene, and CDR3
18 amino acid sequences. Any exact matches between the 2 lists, in which the MIRA TCR sequence
19 was found in at least 2 separate individuals, were considered sufficient to map the ES to the
20 MIRA antigen.

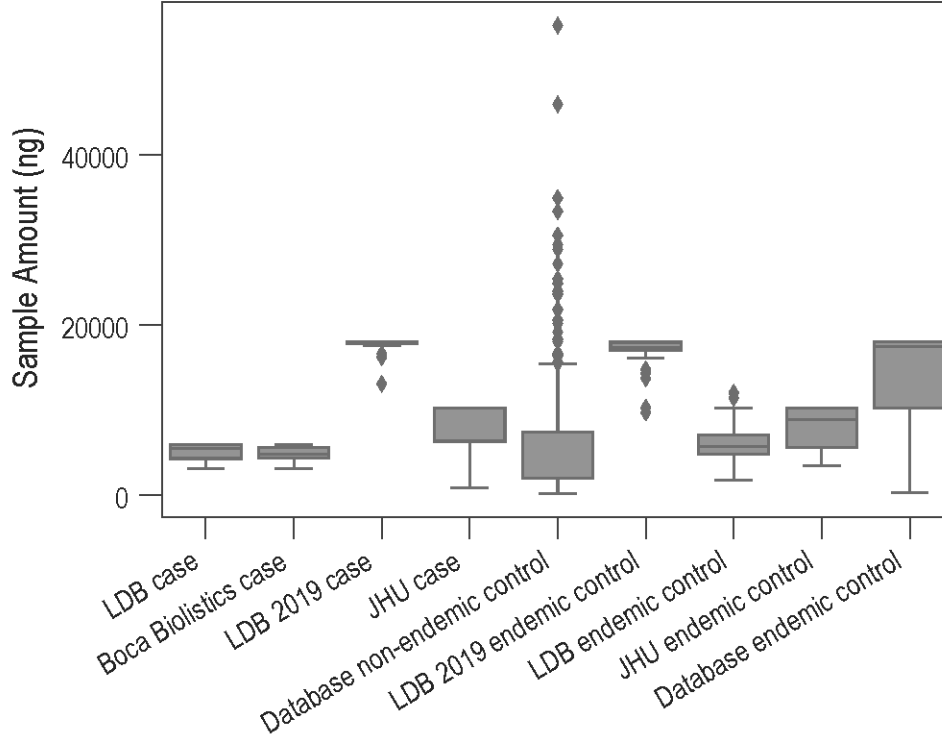
21

1 **ES clustering and HLA inference**

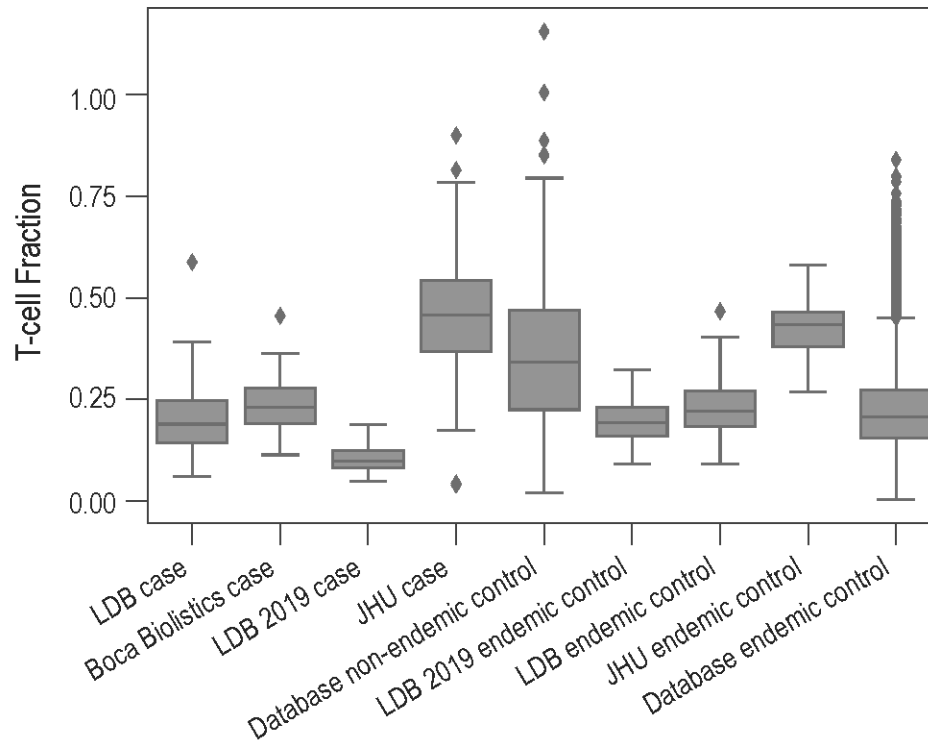
2 Clustering of ESs was based on TCR amino acid similarity. Specifically, 2 TCRs were assigned
3 to the same cluster if they shared a V-gene family (and so have similar CDR1 and CDR2), had
4 identical length, and differed by, at most, 1 amino acid in the CDR3 region. Clusters with at least
5 5 ESs were reported (Table 4). A sequence motif representing the CDR3 amino acid sequences
6 assigned to each cluster was generated using WebLogo (University of California, Berkeley, CA,
7 USA) (Supplemental references 3, 4).

8 To assign an ES to a single HLA subtype, a one-tailed FET was performed between that ES and
9 every HLA subtype. The ES was assigned to the HLA subtype with the lowest P value; if the
10 lowest P value was >0.001 , no assignment was made. Contingency tables counted the number of
11 individuals with/without the ES and with/without a given HLA subtype. For HLA-DQ and HLA-
12 DP, α/β heterodimers were treated as distinct HLA subtypes; for example, individuals with 2 α
13 subtypes and 2 β subtypes were treated as expressing all 4 possible heterodimers. An HLA
14 subtype was assigned to an ES cluster if most ($>50\%$) of the cluster members with an assigned
15 HLA subtype were assigned to the same subtype.

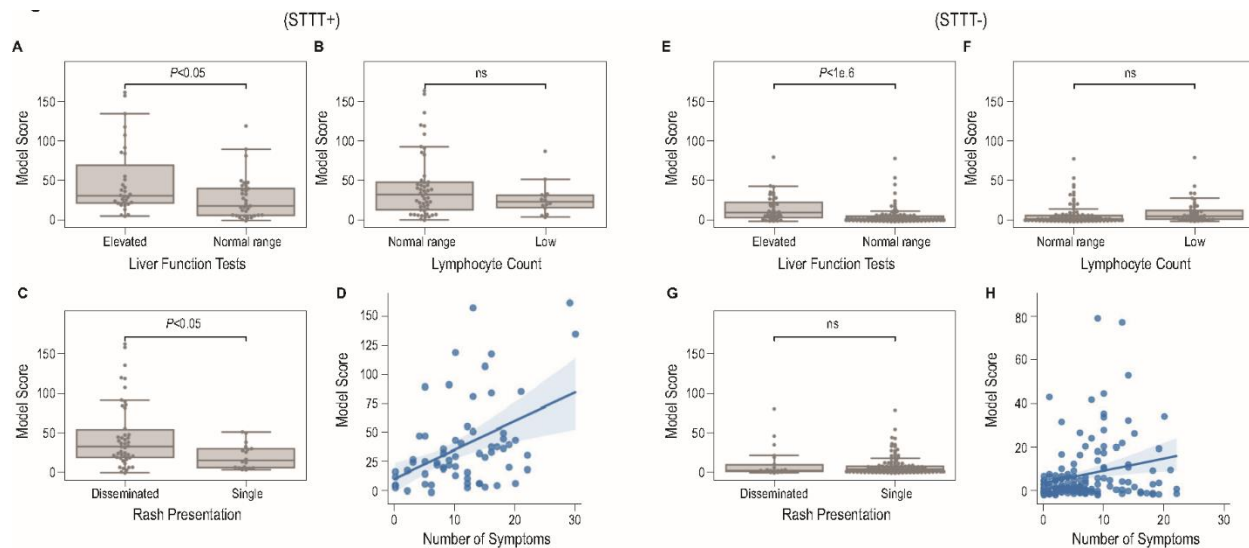
A



B

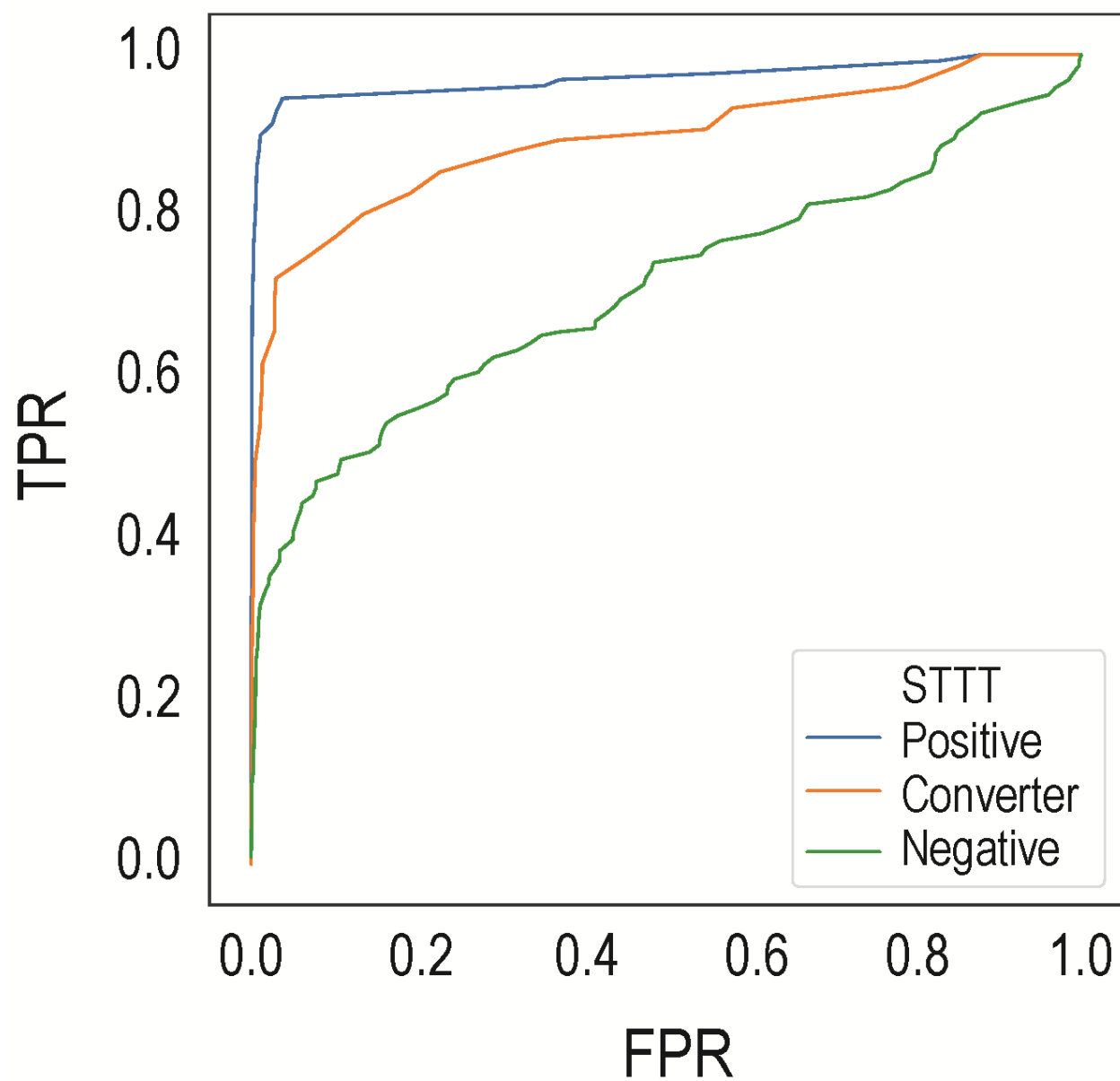


- 1 **Figure S1. Immunosequencing input DNA distributions by cohort.** Boxes indicate median \pm
- 2 IQR, and whiskers denote 1.5 times the IQR above the high quartile and below the low quartile.
- 3



1
2 **Figure S2. Clinical correlates of TCR scoring in STTT-positive (A–D) and STTT-negative**
3 **(E–H) individuals.** TCR scores were stratified by (A, E) liver function test results (elevated
4 [n=72] vs normal [n=139]), (B, F) lymphocyte counts (normal [n=150] vs low [n=61]), or (C, G)
5 presentation of rash (multiple/disseminated [n=68] vs single [n=143]) and (D, H) plotted as a
6 function of the number of Lyme-related symptoms (Spearman $R^2=0.17$. *P* values, Mann-Whitney
7 *U* test). Ns, not significant. In box-and-whisker plots, boxes indicate median \pm IQR, and
8 whiskers denote 1.5 times the IQR above the high quartile and below the low quartile.

9
10
11



1
2
3
4
5
6
7

Figure S3. Receiver operating characteristic curves using all control samples from Lyme-endemic regions from Figure 2A as negatives. Areas under the receiver operating characteristic curves are 0.98, 0.89, and 0.71 for the positive, converter, and negative curves, respectively. FPR, false positive-rate; TPR, true-positive rate.

1 **Table S1. 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
Local/disseminated rash (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

2

1 **Table S2. Counts of enhanced sequences mapped to each protein by MIRA.**
2

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

3
4
5
6
7
8
9
10
11

1 **References cited in the Supplementary Materials only**

- 2
- 3 1. Nolan S, Vignali M, Klinger M, Dines J, Kaplan I, Svejnoha E, Craft T, Boland K,
- 4 Pesesky M, Gittelman R, Snyder T, Gooley C, Semprini S, Cerchione C, Mazza M,
- 5 Delmonte O, Dobbs K, Carreño-Tarragona G, Barrio S, Sambri V, Martinelli G, Goldman
- 6 J, Heath J, Notarangelo L, Carlson J, Martinez-Lopez J, Robins H. 2020. A large-scale
- 7 database of T-cell receptor beta (TCR β) sequences and binding associations from natural
- 8 and synthetic exposure to SARS-CoV-2. Res Sq <https://doi.org/10.21203/rs.3.rs-51964/v1>.
- 9 2. Singh P, Verma D, Backstedt BT, Kaur S, Kumar M, Smith AA, Sharma K, Yang X,
- 10 Azevedo JF, Gomes-Solecki M, Buyuktanir O, Pal U. 2017. *Borrelia burgdorferi* BBI39
- 11 Paralogs, Targets of Protective Immunity, Reduce Pathogen Persistence Either in Hosts or
- 12 in the Vector. J Infect Dis 215:1000.
- 13 3. Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo
- 14 generator. Genome Res 14:1188–1190.
- 15 4. Schneider T, Stephens R. 1990. Sequence logos: a new way to display consensus
- 16 sequences. Nucleic Acids Res 18:6097–6100.