

## RESEARCH ARTICLE

# Comparison of motif-based and whole-unique-sequence-based analyses of phage display library datasets generated by biopanning of anti-*Borrelia burgdorferi* immune sera

Yurij Ionov<sup>1\*</sup>, Artem S. Rogovsky<sup>2\*</sup>

**1** Department of Cancer Genetics, Roswell Park Comprehensive Cancer Center, Buffalo, New York, United States of America, **2** Department of Veterinary Pathobiology, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas, United States of America

\* [Yurij.Ionov@RoswellPark.org](mailto:Yurij.Ionov@RoswellPark.org) (YI); [arogovsky@tamu.edu](mailto:arogovsky@tamu.edu) (ASR)



## OPEN ACCESS

**Citation:** Ionov Y, Rogovsky AS (2020) Comparison of motif-based and whole-unique-sequence-based analyses of phage display library datasets generated by biopanning of anti-*Borrelia burgdorferi* immune sera. PLoS ONE 15(1): e0226378. <https://doi.org/10.1371/journal.pone.0226378>

**Editor:** Nicholas J Mantis, New York State Department of Health, UNITED STATES

**Received:** July 30, 2019

**Accepted:** November 25, 2019

**Published:** January 15, 2020

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0226378>

**Copyright:** © 2020 Ionov, Rogovsky. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

## Abstract

Detection of protection-associated epitopes via reverse vaccinology is the first step for development of subunit vaccines against microbial pathogens. Mapping subunit vaccine targets requires high throughput methods, which would allow delineation of epitopes recognized by protective antibodies on a large scale. Phage displayed random peptide library coupled to Next Generation Sequencing (PDRPL/NGS) is the universal platform that enables high-yield identification of peptides that mimic epitopes (mimotopes). Despite being unsurpassed as a tool for discovery of polyclonal serum mimotopes, the PDRPL/NGS is far inferior as a quantitative method of immune response. Difficult-to-control fluctuations in amounts of antibody-bound phages after rounds of selection and amplification diminish the quantitative capacity of the PDRPL/NGS. In an attempt to improve the accuracy of the PDRPL/NGS method, we compared the discriminating capacity of two approaches for PDRPL/NGS data analysis. The whole-unique-sequence-based analysis (WUSA) involved generation of 7-mer peptide profiles and comparison of the numbers of sequencing reads for unique peptide sequences between serum samples. The motif-based analysis (MA) included identification of 4-mer consensus motifs unifying unique 7-mer sequences and comparison of motifs between serum samples. The motif comparison was based not on the numbers of sequencing reads, but on the numbers of distinct 7-mers constituting the motifs. Our PDRPL/NGS datasets generated from biopanning of protective and non-protective anti-*Borrelia burgdorferi* sera of New Zealand rabbits were used to contrast the two approaches. As a result, the principle component analyses (PCA) showed that the discriminating powers of the WUSA and MA were similar. In contrast, the unsupervised hierarchical clustering obtained via the MA classified the preimmune, non-protective, and protective sera better than the WUSA-based clustering. Also, a total number of discriminating motifs was higher than that of discriminating 7-mers. In sum, our results indicate that MA approach improves the accuracy and quantitative capacity of the PDRPL/NGS method.

**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** BETA, Visualizing clustering of multivariate data; IACUC, Institutional Animal Care and Use Committee of Texas A&M University; NGS, Next Generation Sequencing technology; NZW, New Zealand White rabbits; PCA, Principal Component Analysis; PDRPL, Random peptide phage display library.

## Introduction

Identification of protection-associated (PA) antigens and/or their epitopes via reverse vaccinology is one of the very first steps for development of vaccines against human pathogens [1–9]. To select the most promising vaccine candidate(s) initially requires mapping multiple PA targets, which then can be individually tested for their protective efficacy via animal immunization assays [10]. To map numerous PA epitopes, high throughput methods, which would allow PA targets to be specifically recognized by protective serum antibodies on a large scale, are needed.

The high-density peptide microarray is a traditional method for mapping linear epitopes of antibodies developed against protein antigens [11–13]. The microarray contains overlapping peptides that cover the entire length of a target protein. However, this relatively straightforward method is not optimal for analyzing repertoires of antibodies developed against intact bacterial pathogens because these microorganisms are represented by an enormously complex mixture of antigens. Generation and application of customized microarrays with peptides that would encompass an entire proteome of a virus or bacterium is highly cost-prohibitive and time-consuming.

An alternative to customized platforms is the random peptide array, where  $10^4$  17-mer random peptide sequences selected by a random number generator are printed on glass slides [14]. When a serum sample is applied to the surface of this peptide array, distinct binding patterns of the unique molecular recognition elements associated with pathogen-specific antibodies are generated. The random peptide array is highly reproducible and, hence, allows changes in the antibody repertoires that are characteristic for different states of a disease to be detected [15, 16]. However, the disadvantage is that densely surface-immobilized ligands can amplify non-specific interactions of low affinity, which reduces the overall capacity of the random peptide array for identifying disease-specific interactions. The other drawback is that the total number of peptides that can be attached to a single 25×75-mm glass is limited. Despite previous studies successfully used  $10^4$  random peptides to characterize different immune responses [17], the assay with such a low number of ligands can still miss important antibody reactivities.

Phage displayed random peptide library (referred to here as PDRPL) is an inexpensive alternative to the high-density peptide array. The first application of PDRPL demonstrated the feasibility of affinity selection for identifying a peptide mimetic, which represented cognate epitopes of monoclonal antibodies [18]. Since then, PDRPL has been extensively applied to characterize specificity of polyclonal sera from cancer or infectious disease patients [19–24]. Similar to the random peptide array, PDRPL is also a universal platform, yet with a much higher capacity to represent sequences of a complete proteome of any microorganism. For example, a commercially available Ph.D.-7 library of random heptapeptides (New England Biolabs, USA) has a complexity on the order of  $10^9$  independent clones and thus can represent any unique 7-mer peptide sequence (referred to here as 7-mer sequence). According to the product specifications, the phage concentration of the Ph.D.-7 library is  $10^{13}$  phage particles per mL. The latter means that 10  $\mu$ L of the library used for a biopanning experiment contains  $10^{11}$  phage particles and each 7-mer is thus represented on average by almost 80 times (which is calculated by dividing  $10^{11}$  by  $20^7$ ). In addition to linear epitopes, phage displayed peptides may also represent both linear and conformational epitope mimetics of protein and carbohydrate antigens also known as mimotopes [25].

Until a recent advent of Next Generation Sequencing technology (referred to here as NGS), the main disadvantage of traditional PDRPL was the necessity to individually sequence phage clones. The NGS has transformed the phage display into a high throughput method, which, to date, allows the diversity of antibody specificities to be thoroughly examined [26]. By using the

NGS, hundred thousands of affinity-selected peptide sequences can be obtained in a single sequencing run [27]. As a result, the PDRPL coupled with the NGS (referred to here as PDRPL/NGS) has been extensively utilized to analyze repertoires of serum antibodies [28–30].

In our recent studies [10, 31, 32], we successfully applied the PDRPL/NGS to define repertoires of antibodies developed in mice and rabbits that had an active infection with *Borrelia burgdorferi*, the bacterial agent of Lyme disease (LD) [33–41]. LD pathogen has the capacity to establish persistent infection in mice, the primary natural mammalian reservoir of *B. burgdorferi* in the United States [42–47], and human patients [48–53], despite very strong anti-*B. burgdorferi* antibody responses [54–60]. In immunocompetent mouse models, the ability of *B. burgdorferi* to successfully evade otherwise potent antibodies is mainly attributed to the highly efficacious VlsE-encoding system, whose genetic removal results in rapid clearance of LD pathogen by host antibodies [54, 59, 61–69]. In contrast to experimental mouse or any other animal LD models [70–86], however, New Zealand White (NZW) rabbits develop a protective antibody-mediated response, which effectively clears *B. burgdorferi* within 3–8 weeks, despite the antigenically varying surface protein, VlsE [32, 87, 88]. Our recent study has examined repertoires of antibodies in sera collected from *B. burgdorferi*-infected NZW rabbits at day 14 and 28 postinfection, the time points at which the rabbit antibodies were shown to be non-protective or protective against the LD pathogen in mice, respectively [32].

In the present work, we used our previously generated PDRPL/NGS data to improve on the PDRPL/NGS analysis in discriminating between the protective and non-protective anti-*B. burgdorferi* sera of NZW rabbits [32]. Specifically, in order to extract more information on any qualitative and quantitative alterations in anti-*B. burgdorferi* antibody specificities, we have compared two distinct modes of analyses of the PDRPL/NGS data. The whole-unique-sequence-based analysis (referred to here as WUSA) involved 7-mer sequences and compared the number of sequencing reads for each unique peptide between the protective and non-protective sera after the normalization for the total number of all the sequences. In contrast, the alternative motif-based (referred to here as MA) method included identification and comparison of four-amino-acid-long consensus motifs (referred to here as 4-mer motifs), each representing a cluster of related 7-mer sequences that differed by only few amino acids.

## Results

Our previous study has demonstrated that by day 28 postinfection, NZW rabbits developed a repertoire of protective antibodies against *B. burgdorferi* infection [32]. Specifically, it was shown that, when passively transferred to naïve mice, the rabbit antibodies prevented infection by highly immune-evasive *B. burgdorferi*. Moreover, in mice with ongoing infection, the rabbit antibodies were able to significantly reduce LD-induced inflammation in joints [32], the avascular collagenous tissues to which antibody access is somewhat limited [89]. Our follow-up study defined specificities of day-14 (non-protective) and day-28 (protective) rabbit antibodies via the PDRPL/NGS and then directly compared the two repertoires. As a result, specificities of the rabbit protective antibodies and their respective targets were identified [9].

In the present study, we have compared two distinct approaches, the WUSA and MA, for the analysis of the PDRPL/NGS data [9]. Both approaches directly compared mimotope sequences between the day-14 (non-protective) and day-28 (protective) immune serum samples from 3 NZW rabbits (animals P, Y, and Z). The preimmune sera were pooled from the 3 rabbits prior to their challenge with wild-type *B. burgdorferi* B31-A3 strain and served as a background control. To obtain mimotope sequences, 1 pooled preimmune, 3 non-protective, and 3 protective serum samples (referred to here as PI, NP, and PR samples, respectively) were analyzed via commercially available Ph.D.-7 library of random heptapeptides followed by

sequencing via Illumina 2500 sequencing platform [9]. One PR sample was analyzed in duplicate (samples P28a and P28b). For both methods, 1,000 most abundant 7-mer sequences generated from the PDRPL/NGS analysis of NP (P14, Y14, and Z14), PR (P28a, P28b, Y28, and Z28), and PI samples were selected.

### Analyzing the protective and non-protective rabbit serum samples via the MA approach

In the MA approach, we first identified mimotope motifs that represented clusters of 7-mer sequences. Within each cluster, the respective 7-mers differed by few amino acids outside their 4-mer motif. Motifs, in turn, could have conservative amino acid substitutions. To generate 50 4-mer motifs for each sample, a total of 8 mimotope sequence datasets (P14, Y14, Z14, P28a, P28b, Y28, Z28, and PI) were separately processed via MEME software tool (<http://meme-suite.org/tools/meme>) [90]. Each dataset consisted of 1,000 7-mers and each motif was derived from at least 3 distinct sequence variants. To make the 7-mer sequences compatible with the MEME tool (the minimum length is 8 letters), the letter X was added to each 7-mer. Prior to running the program, we chose that each motif has the minimum sites of 3 (to unify at least 3 unique sequences) and the width of 4 amino acids. The MEME output exemplified in Fig 1 shows two abundant motifs, QKPL and KIGD, which unified 9 and 46 distinct 7-mer sequences, respectively, for sample Z28. As a result of this *in silico* analysis, we found that

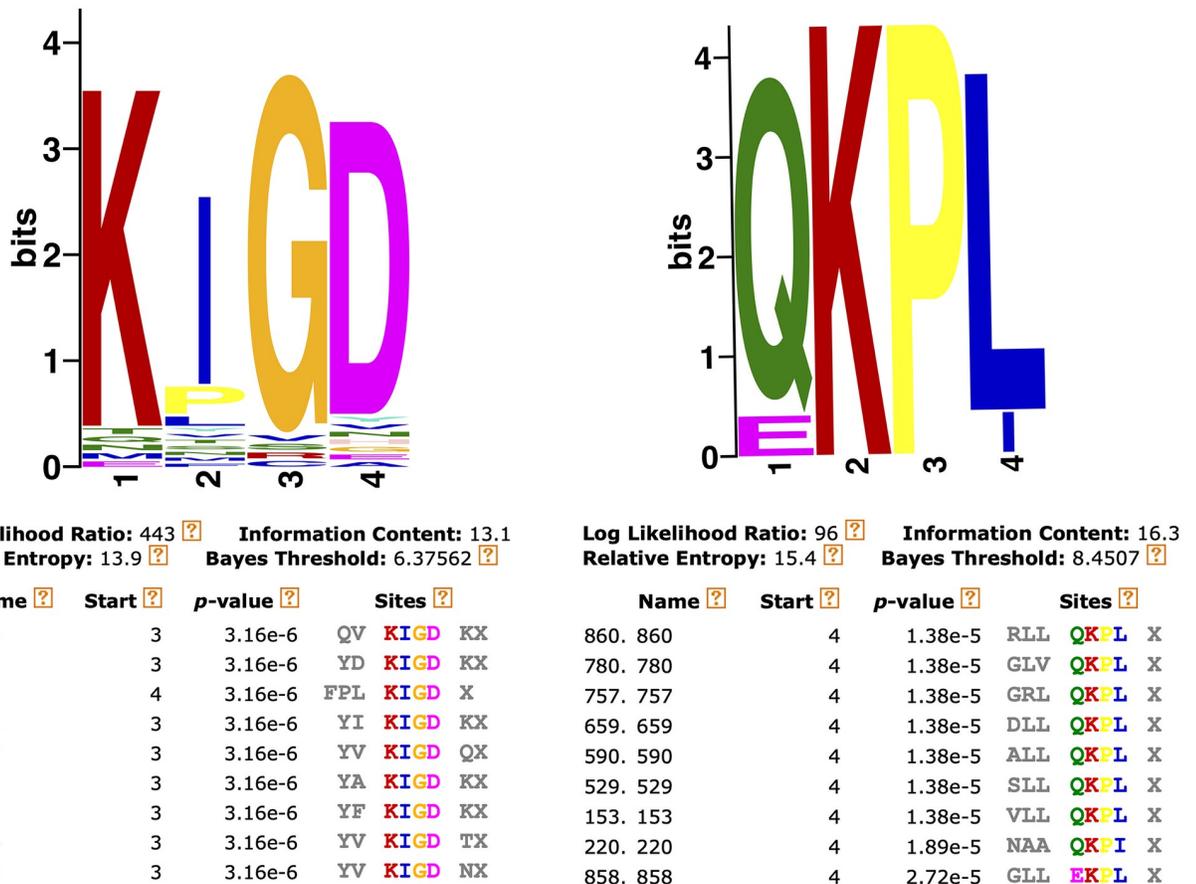
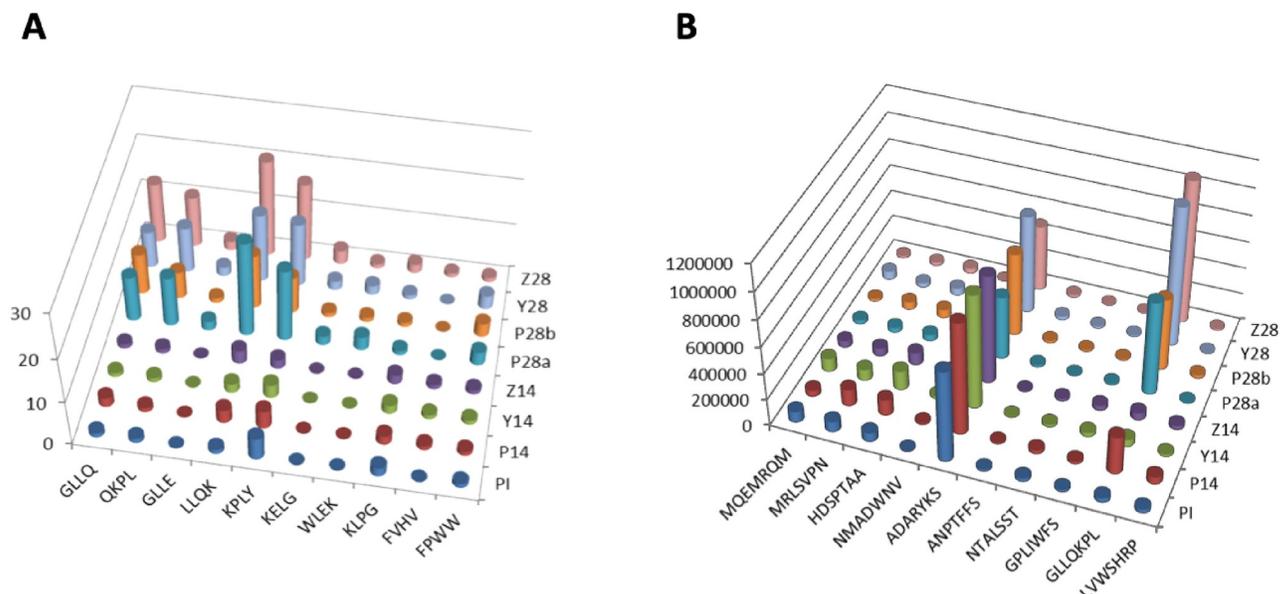


Fig 1. Examples of MEME-generated 4-mer motifs unifying the unique 7-mer peptide sequences.

<https://doi.org/10.1371/journal.pone.0226378.g001>

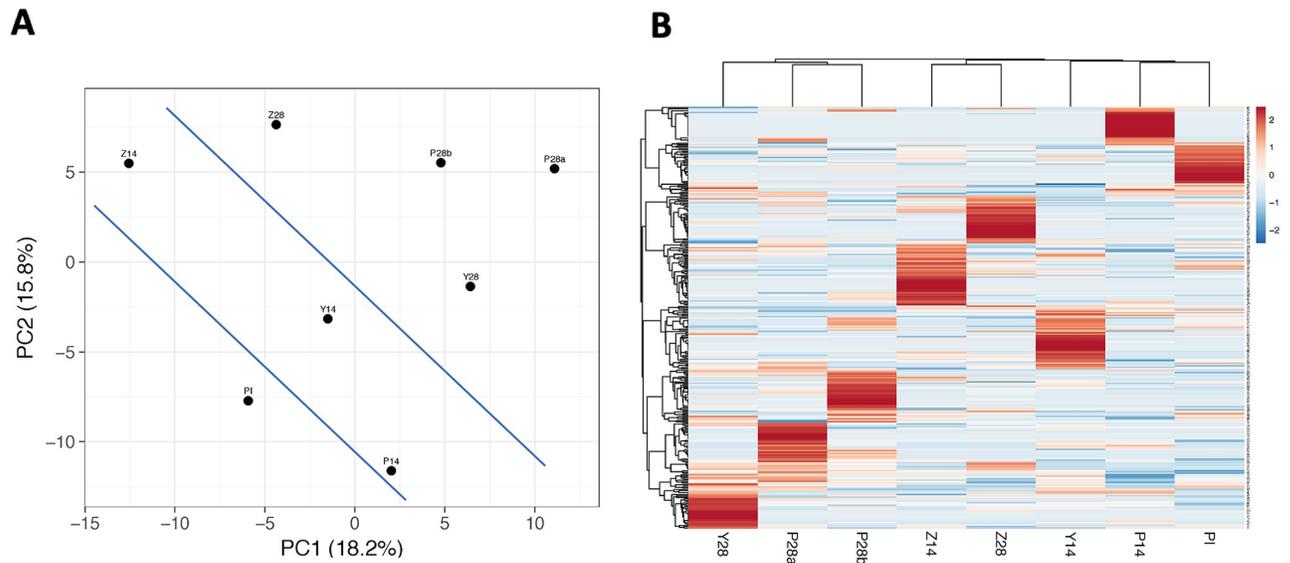


**Fig 2. Frequency distribution across rabbit serum samples for the discriminating 4-mer motifs (A) and unique 7-mer peptides (B) with the lowest  $p$ -values according to the  $t$ -test.**

<https://doi.org/10.1371/journal.pone.0226378.g002>

some motifs were shared by the 8 datasets. After the redundant motifs were removed, there was a total of 330 distinct motifs for the 8 samples. [S1 Table](#) provides a complete list of these motifs and the respective numbers of the 7-mer sequences that constituted each motif for each serum sample. After performing the paired two-tailed  $t$ -test (leaving out one of the two technical replicates for one protective sample), we identified 9 most significant motifs that discriminated between the NP and PR samples ( $p < 0.1$ ). Furthermore, out of these 9 motifs, 4 motifs were most significant ( $p < 0.05$ ) and were derived from a most frequently represented 7-mer for the 4 PR samples, GLLQKPL. [Fig 2A](#) shows the distribution of 10 selected motifs that were most significant for the 8 samples. Since the  $t$ -test may be not the best statistics to identify discriminating motifs, a different method was also tested. For that, the minimum number of motif variants for the PR samples was compared with the maximum number of motif variants for the NP samples. As a result, we identified a total of 10 motifs, whose minimum numbers for the PR samples were higher than the maximum numbers for the NP samples. Interestingly, 7 of these 10 motifs were identical to the most statistically significant motifs identified via the  $t$ -test ([S1 Table](#), Sheet 2).

We then performed Principal Component Analysis (PCA) on the 330 motifs by using the web tool, BETA, which was developed to visualize clustering of multivariate data (<https://biit.cs.ut.ee/clustvis/>) [91]. Since the first principle component, which determines the direction of highest variability in the data, captured only 18% of the variability, the plot showed a high degree of data randomness ([Fig 3A](#)). Despite the observed randomness, however, the PCA analysis still separated the 4 PR samples from all the 3 NP and 1 PI samples ([Fig 3A](#)). The heatmap generated by the unsupervised hierarchical clustering via the ClustVis also demonstrated a high degree of variability between the 8 samples ([Fig 3B](#)). However, similar to the PCA plot, the respective heatmap showed that the PI sample was well segregated from the 4 PR samples; and yet neighbored with a cluster of the 3 NP samples ([Fig 3B](#)). Consistently, the duplicate samples, P28a and P28b, were grouped together despite a high degree of variability within each technical replicate ([Fig 3B](#)). Of note, the proximity of PCA values for samples Z14 and Z28 and



**Fig 3. Principal Component Analysis (PCA) plot (A) and heatmap (B) generated, via the ClustVis, by the motif-based analysis (MA).** PDRPL/NGS data were produced by biopanning of preimmune (PI), non-protective (P14, Y14, and Z14), and protective (P28a, P28b, Y28, and Z28) rabbit serum samples. Blue lines drawn across the PCA plot (A) separate the 3 groups of serum samples.

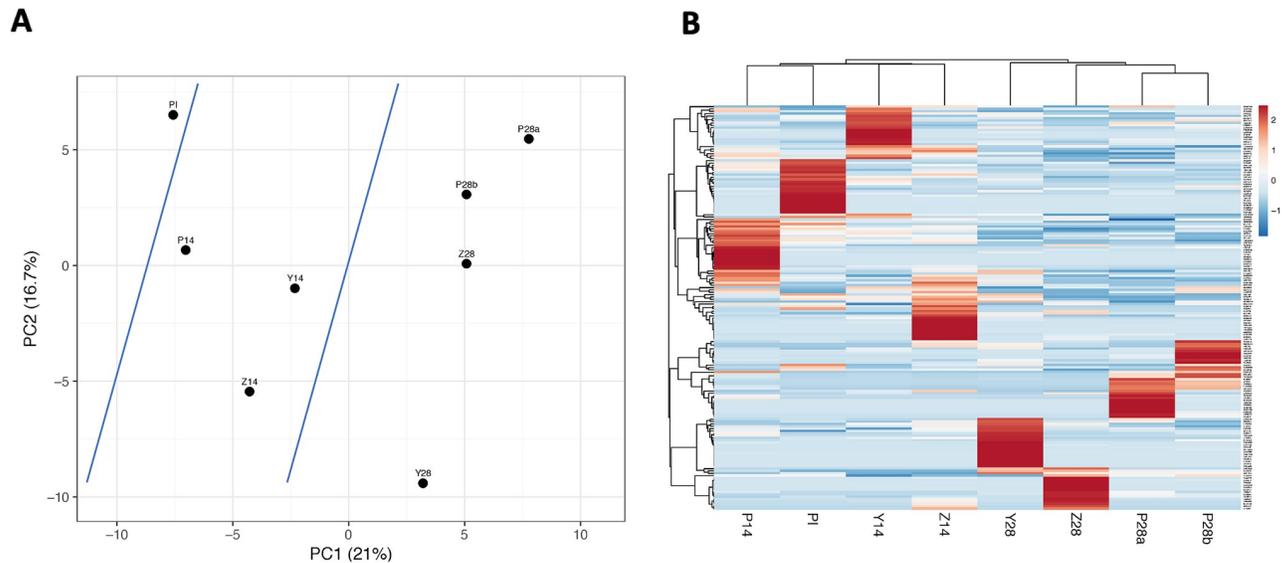
<https://doi.org/10.1371/journal.pone.0226378.g003>

their respective heatmap patterns suggested that NZW rabbit Z had developed a protective antibody repertoire later compared to animals P and Y. Indeed, in contrast to NZW rabbits P and Y, skin biopsies of animal Z sampled at day 28 postinfection were only weakly culture-positive for *B. burgdorferi*, whose very low spirochetal numbers and impaired (slow) motility indicated a relatively late clearance [32].

### Analyzing the protective and non-protective rabbit serum samples via the WUSA approach

To compare the discriminating capacity of the MA with that of the more conventional WUSA method, the same 8 mimotope sequence datasets (P14, Y14, Z14, P28a, P28b, Y28, Z28, and PI) were analyzed via the WUSA. First, we calculated the number of sequencing reads for each 7-mer sequence, which reflected the relative titers of the respective mimotope-recognizing antibodies (1S Table). Thus, the numbers of sequencing reads were used to directly compare levels of antibodies of distinct specificities within and between the serum samples. Importantly, the numbers of 7-mers for each sample were normalized to have totals of sequencing reads equal across the 8 samples. This normalization was necessary to eliminate any biases that are associated with potential differences in amplification efficiency between the PCR reactions performed prior to multiplexing a phage DNA library.

For each dataset, we selected the top 50 most abundant sequences to match with the number of motifs ( $n = 50$ ) used for the MA method. After eliminating the redundant sequences, there was a total of 172 distinct 7-mers for the 8 datasets. S2 Table lists the unique 7-mer sequences and the respective numbers of their sequencing reads. After performing the paired two-tailed  $t$ -test, we identified 11 most significant 7-mers that discriminated between the NP and PR samples ( $p < 0.1$ ). Fig 2B shows the 10 peptide sequences that had the lowest  $p$ -values. Then, the minimum numbers of sequencing reads for the PR samples were compared with the respective maximum values for the NP samples. Consequently, only 3 out of 11 peptides ( $\alpha = 0.1$ ) had the higher numbers of their sequencing reads for the PR samples compared to the NP



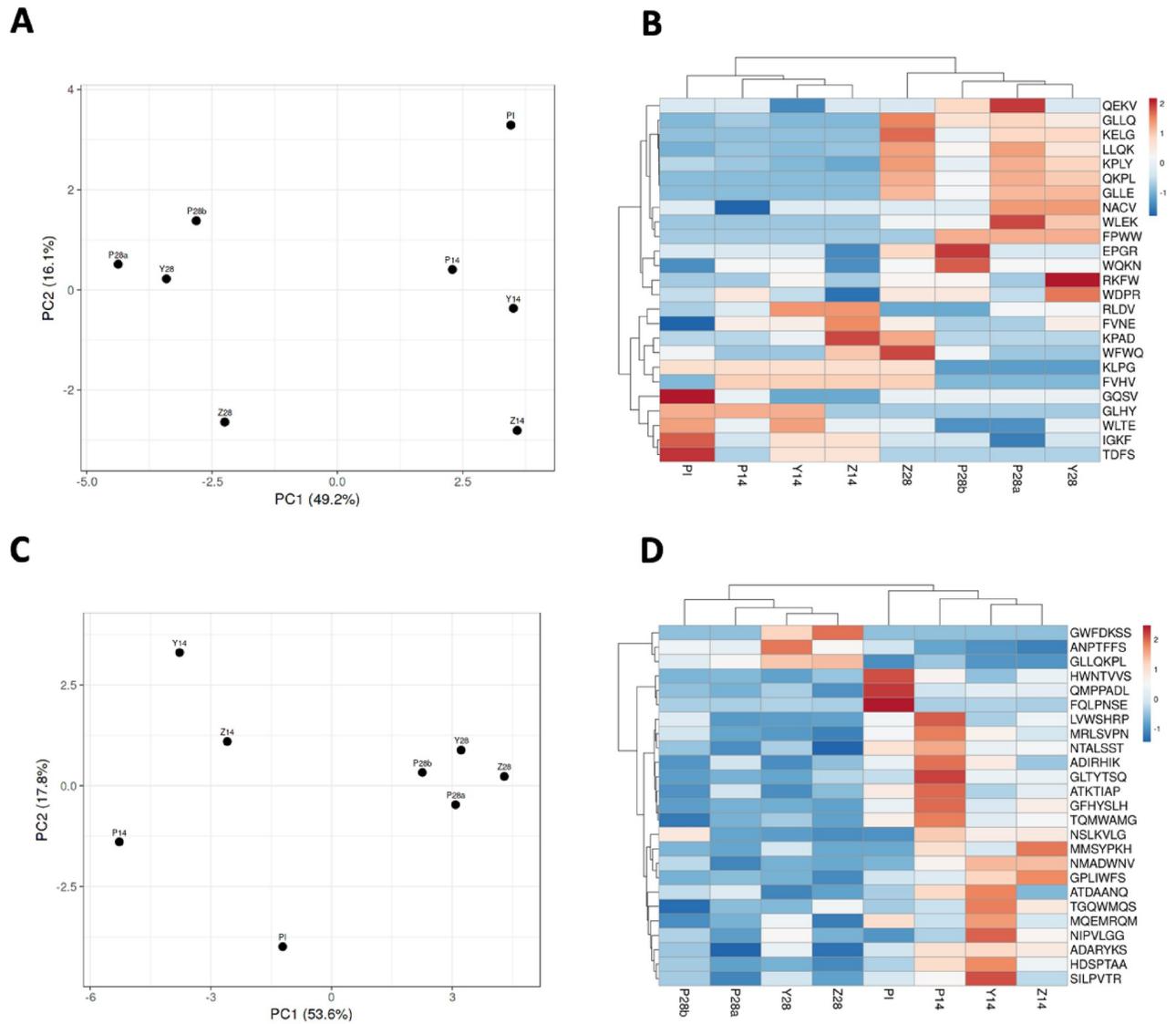
**Fig 4. Principal Component Analysis plot (A) and heatmap (B) generated, via the ClustVis, by the whole-unique-sequence-based analysis (WUSA).** PDRPL/NGS data were produced by biopanning of preimmune (PI), non-protective (P14, Y14, and Z14), and protective (P28a, P28b, Y28, and Z28) rabbit serum samples. Blue lines drawn across the PCA plot (A) separate the 3 groups of serum samples.

<https://doi.org/10.1371/journal.pone.0226378.g004>

samples. Out of these 3 peptides, GLLQKPL was most discriminating as it had the highest positive difference in the number of its sequencing reads between the PR and NP samples (Fig 2B). Consistently, the PCA analysis showed a high degree of data randomness with the first principle component capturing only 21% of the variability. Despite this randomness, the PI, NP, and PR samples were still segregated by the PCA plots (Fig 4A). Interestingly, the respective heatmap did not discriminate between the PI and NP samples (Fig 4B). Consistent with the MA method results, the duplicate samples, P28a and P28b, were also clustered together; although this time this overlap was slightly higher (Fig 4B).

### Improving on the MA and WUSA methods

The high degree of data randomness shown by both WUSA and MA methods may be due to a high level of noise inherently associated with the PDRPL/NGS. In addition, the observed randomness could indicate that only a small fraction of antibody repertoires of infected NZW rabbits is related to an anti-*B. burgdorferi* immune response; whereas the majority of rabbit antibody repertoires may simply reflect the previous exposure to unrelated antigens. To increase the signal to noise ratio, this time we generated the PCA plots and respective heatmaps only for the first 25 motifs and the first 25 7-mer sequences as ranked by the *t*-test. The PCA plots showed that the overall separation between the PR and NP samples were noticeably improved by both WUSA and MA methods (Fig 5A and 5C). Overall, the heatmap generated via the MA method reflected the immunological status (protective vs. non-protective responses) more accurately than the WUSA-based heatmap (Fig 5B and 5D). Curiously, the arrangement of the 8 samples from left to right showed the progression of protective immune response against the LD pathogen. Furthermore, sample Z28 from the NZW rabbit that had developed its protective antibodies later than the other animals was placed on the border between the PR and NP samples and was adjacent to sample Z14 derived from the same rabbit (Fig 5B). The replicates, P28a and P28b, were clustered together and the PI sample was distantly positioned from the 4 PR samples (Fig 5B). On the contrary, the heatmap generated via



**Fig 5. ClustVis-generated Principal Component Analysis plots (A and C) and heatmaps (B and D) produced for the first 25 most statistically significant 4-mer motifs (A and B) and unique 7-mer sequences (C and D).** PDRPL/NGS data were generated by biopanning of preimmune (PI), non-protective (P14, Y14, and Z14), and protective (P28a, P28b, Y28, and Z28) rabbit serum samples.

<https://doi.org/10.1371/journal.pone.0226378.g005>

the WUSA, placed the PI sample between the PR and NP samples; and yet the technical replicates, P28a and P28b, were consistently grouped together (Fig 5D).

### Discussion

Reverse vaccinology is a powerful vaccine development approach, which involves initial identification of vaccine targets via genome-based computational analysis [92]. In our recent studies, we utilized a subtractive reverse vaccinology to delineate putative surface epitopes of *B. burgdorferi* associated with antibody-mediated protection against the LD pathogen [10, 32, 93]. By using the two LD animal models, protective and non-protective immune sera from C3H mice and NZW rabbits were analyzed via the PDRPL/NGS and compared within each animal species [10, 93]. To further evaluate the utility of the PDRPL/NGS for quantitative analysis of

immune responses, in the present study, we conveniently made use of the PDRPL/NGS datasets previously generated from the NZW rabbit model [93].

Although the PDRPL/NGS is a very sensitive high throughput method for identifying disease-specific mimotopes recognized by serum antibodies [94]; its reproducibility and accuracy are low [30] compared to the random peptide array [14, 15, 95]. There are at least two probable causes for the low reproducibility. First, the binding of low abundant antibodies to their cognate targets is a stochastic process, since, during the first round of selection, these targets are present in very low numbers. Second, there is a loss of relevant sequences over secondary rounds of selection and propagation of phages [30]. Importantly, these multiple rounds of selection and amplification are required for making the ratio of a target to its cognate antibody large enough to reach the positive correlation between the number of sequencing reads for a given unique peptide and a titer of its specific antibody. However, because of these required secondary rounds, the phage amplification becomes highly sensitive to fluctuations in the number of bacterial cells successfully infected by competing phage particles. As a result, the numbers of sequencing reads in technical replicates may show a large variation, which in turn reduces the overall accuracy of the quantitative analysis of the PDRPL/NGS data.

To improve the reproducibility of the PDRPL/NGS method, in the present study, we evaluated the feasibility of using 4-mer motifs as potential biomarkers of the protective anti-*B. burgdorferi* immune response. Given that most of the binding energy is typically derived from 4–6 amino acids [17], the specificity of antigen-antibody interaction is defined by only a few anchor residues within an epitope [29, 96–98]. Previously, we [27] and others [34] demonstrated that the four-amino-acid consensus is a minimum motif length, which is sufficient to define an antibody specificity. Thus, the high copy number of shorter targets in the initial selection makes the process of target recognition more deterministic. While the unique 7-mer sequence is represented on average by almost 80 copies, the tetramer is represented by a much greater number of copies:  $10^{11}/20^4 = 625,000$ . Moreover, it was previously noted that the size of a peptide family (a total number of sequence-related peptides) sharing a given motif positively correlated with a number of sequencing reads for the most represented sequence of that family and the titer of serum antibodies recognizing that motif [29]. Importantly, this positive correlation was experimentally confirmed via ELISA by utilizing immobilized motif-containing peptides [29]. Thus, considering the numbers of distinct motif-constituting sequences transforms qualitative peptide differences into a quantitative parameter, which more accurately reflects titers of target-specific antibodies.

The present study showed that, in general, the MA method discriminated between the protective and non-protective serum samples as equally well as the WUSA with a couple of noteworthy differences. First, although the number of the discriminating 7-mer sequences identified by the *t*-test ( $\alpha = 0.1$ ) is a little higher than that of discriminating motifs (11 vs. 9), only 3 out of 11 whole sequence peptides were represented by higher numbers of their respective sequencing reads for the PR samples compared to the NP samples. In contrast, 7 out of 9 motifs well discriminated the PR samples ( $p = 0.024619$  by Chi-Square test). Second, the unsupervised hierarchical clustering obtained via the MA method arranged the samples in the order that more accurately reflected the development of the protective anti-*B. burgdorferi* immune response compared to the WUSA-based clustering. The observed outcome difference can be attributed to the fact that the MA method is more quantitative and, hence, more discriminating by nature compared to the WUSA approach. For example, when group A and group B sera (3 samples in each group) are compared and each of the 3 samples of group A has only 1 unique 7-mer; and together these 7-mers form a motif, then it is only the motif (rather than the 7-mers) that will discriminate between the two groups. Also, the MA method tends to level off high fluctuations in numbers of sequencing reads for 7-mer sequences. For example,

despite the GLEKLH peptide was consistently detected and absent in all the PR and NP samples, respectively; the high variance in its sequencing reads prevented this 7-mer from distinguishing between the two serum groups. To contrast, the GLE motif shared by this and other 7-mers of the 4 PR samples was well discriminating. Therefore, the MA approach allows each of motif-constituting 7-mers to quantitatively contribute to the overall discriminatory power of the respective motif regardless of high variations in its sequencing reads. Given that a total number of a given unique peptide positively correlates with the number of its respective sequencing reads [29]; motifs, in general, are likely to more accurately reflect both quantitative and qualitative alterations in antibody repertoires.

## Methods

### Bacteria and rabbit infection

In the prior study [32], *B. burgdorferi* B31-A3 strain [99] was utilized to infect New Zealand White (NZW) rabbits. Spirochetes were cultivated in liquid Barbour-Stoenner-Kelly medium supplemented with 6% rabbit serum (Gemini Bio-Products, USA; referred to here as BSK-II) and incubated at 35°C under 2% CO<sub>2</sub>. For our prior study [32], a total of 3 female NZW rabbits of 12–14 weeks of age (rabbits P, Y, and Z; 3.0–3.5 kg by weight each) were purchased from Charles River Laboratories (USA). The rabbits were singly housed in an animal BSL-2 facility at Texas A&M University. The animals were daily subjected to visual welfare-related assessment for the entire duration of their stay. In order to obtain preimmune sera, the 3 rabbits were bled via the marginal ear vein after their acclimation. Each animal was then individually inoculated intradermally at six sites along the spine with *B. burgdorferi* B31-A3 strain at 10<sup>6</sup> cells per each inoculation site as described [88]. The infection of each animal was verified by culture-positive skin biopsies taken around inoculation sites (3 mm in diameter) at weeks 1, 2, and 3 postinfection. Skin tissues collected at week 4 postinfection were also collected and showed that all the skin biopsies were culture-negative with an exception being NZW rabbit Z, whose skin culture was weakly positive. In addition to very low numbers of spirochetes in the NZW rabbit Z skin culture, spirochetes exhibited no or very little (impaired) motility, which indicated the final stage of spirochetal clearance [32]. The blood samples were placed at 4°C until next day. Serum samples were then generated from each blood sample via centrifugation at 5,000 x g for 10 minutes. The serum samples were stored at -80°C. To prevent bacterial and fungal contamination during skin biopsy culture, BSK-II was supplemented with 0.02 mg ml<sup>-1</sup> phosphomycin, 0.05 mg ml<sup>-1</sup> rifampicin, and 2.5 mg ml<sup>-1</sup> amphotericin B. All the cultures were weekly examined for 6 weeks via dark-field microscopy for the presence of viable spirochetes. At week 4 postinfection, the NZW 3 rabbits were humanely sacrificed by applying isoflurane.

### Phage display (Ph.D.) library

Twenty µl of each rabbit serum sample and 10 µl of random peptide library Ph.D.-7 (NEB, MA, USA) were diluted in 200 µl of TRIS Buffered Saline buffer with 0.1% Tween 20 (referred to here as TBST) and 1% bovine serum albumin and incubated at 25°C for 18 hr [29]. Antibody-bound phages were isolated by utilizing protein G-agarose beads (Santa Cruz Biotechnology, Inc., TX, USA) to the phage antibody mixture for 1 hr. The bead mixture was transferred to a 96-well MultiScreen-Mesh filter plate (EMD Millipore, MA, USA) with a 20-µm-pore-size nylon mesh at the bottom. To remove unbound phages, vacuum was used to the exterior of the nylon mesh. The beads were washed 4 times with 100 µl of TBST buffer for each well. Antibody-bound phages were eluted with 100 µl of 100 mM Tris-glycine buffer (pH 2.2) and subsequently neutralized by adding 20 µl of 1 M Tris buffer (pH 9.1). The solution was

then used for amplification of eluted phages by infecting bacteria per the manufacturer's instructions. Amplified phages were subjected to the next round of biopanning, after which antibody-bound phages were isolated by using protein G-agarose beads. After washing the beads in the 96-well MultiScreen-Mesh filter plate, DNA was isolated by phenol-chloroform extraction and ethanol precipitation. Lastly, 21-nucleotide-long DNA fragments that encoded random peptides were amplified by PCR utilizing the following forward and reverse primers, respectively: 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGAT CT (INDEX) TGGTACCTTTCTATTCTCACTCT-3' and 5'-CAAGCAGAAGAGGGCATAACGAGCTCTTCCGATCTAACAGTTTCGGCCGAACCTCCACC-3'. The INDEX of each forward primer was defined by the unique six-nucleotide long sequence. Thus, for each serum, a forward primer with the unique INDEX was used [29]. After purification, generated PCR-amplified DNA library was multiplexed and sequenced by 50 single read cycles via Illumina HiSeq 2500 platform at University of Buffalo Genomics and Bioinformatics Core, New York State Center of Excellence Bioinformatics & Life Sciences, Buffalo, New York.

### DNA reads analysis

The sequencing resulted in approximately  $1.79 \times 10^8$  DNA reads. Reads were de-multiplexed based on the unique bar codes. Each read was tagged by the unique INDEX and contained a 21-nucleotide-long sequence, which encoded a given peptide. The 21-nucleotide sequences were extracted between positions 30 and 50 and translated into 7-mers in the first frame. The mean number of peptides that had no stop codon per serum was  $5.9 \times 10^6$ , of which approximately  $2.1 \times 10^5$  peptides per sample were non-redundant. The numbers of peptides were normalized in order that the total number of peptides to be equal between samples and the first 1,000 most abundant unique peptide sequences were selected for each serum samples. For each serum sample, S3 Table shows the first 1,000 unique peptides with the corresponding normalized number of sequencing reads for each unique peptide sequence.

### Generating motifs using MEME software

One thousand most abundant unique peptide sequences were uploaded to MEME software available online (<http://meme-suite.org/tools/meme>). The parameters of running the algorithm were selected to generate 50 motifs. To make the 7-mer sequences compatible with the MEME tool requirements (the minimum length is 8 letters), the letter X was added to each 7-mer. Prior to running the program, we chose that each motif has the minimum sites of 3 (to unify at least 3 unique sequences) and the width of 4 amino acids exactly.

### Declarations

#### Ethics approval and consent to participate

NZW rabbits were maintained at Texas A&M University in an animal facility accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care International. All the experimental practices, which involved animals, had been approved by the Institutional Animal Care and Use Committee of Texas A&M University (IACUC 2017-0390) and were carried out in accordance with Public Health Service Policy on Humane Care and Use of Laboratory Animals (2002), Guide for the Care and Use of Agricultural Animals in Research and Teaching (2010), and Guide for the Care and Use of Laboratory Animals (2011).

## Supporting information

**S1 Table. Sheet 1.** Frequency distribution for the discriminating 4-mer motifs with  $p$ -values calculated by the  $t$ -test to distinguish between the non-protective (P14, Y14, and Z14) and protective (P28, Y28, and Z28) rabbit serum samples. **Sheet 2.** Frequency distribution for the discriminating 4-mer motifs that were ranked by the difference between the minimum values of the protective serum samples (P28, Y28, and Z28) and the maximum values of the non-protective serum samples (P14, Y14, and Z14). Highlighted in yellow are statistically significant motifs.

(XLSX)

**S2 Table. Sheet 1.** Frequency distribution for the discriminating unique 7-mer sequences with  $p$ -values calculated via the  $t$ -test to distinguish between the non-protective (P14, Y14, and Z14) and protective (P28, Y28, and Z28) rabbit serum samples. **Sheet 2.** Frequency distribution for the discriminating unique 7-mer sequences that were ranked by the difference between the minimum values of the protective serum samples (P28, Y28, and Z28) and the maximum values of the non-protective serum samples (P14, Y14, and Z14).

(XLSX)

**S3 Table. The first one thousand of the most abundant 7-mer peptides with the respective numbers of their sequencing reads for each rabbit serum sample.** PI denotes pooled preimmune serum sample. P14, Y14, and Z14 are non-protective serum samples. P281, P28b, Y28, and Z28 are protective serum samples. P28a and P28b are technical replicates.

(XLSX)

## Author Contributions

**Conceptualization:** Yuriy Ionov, Artem S. Rogovskyy.

**Data curation:** Yuriy Ionov, Artem S. Rogovskyy.

**Formal analysis:** Yuriy Ionov, Artem S. Rogovskyy.

**Funding acquisition:** Artem S. Rogovskyy.

**Investigation:** Yuriy Ionov, Artem S. Rogovskyy.

**Methodology:** Yuriy Ionov.

**Project administration:** Artem S. Rogovskyy.

**Resources:** Artem S. Rogovskyy.

**Supervision:** Artem S. Rogovskyy.

**Validation:** Artem S. Rogovskyy.

**Visualization:** Artem S. Rogovskyy.

**Writing – original draft:** Yuriy Ionov.

**Writing – review & editing:** Artem S. Rogovskyy.

## References

1. Pizza M, Scarlato V, Masignani V, Giuliani MM, Arico B, Comanducci M, et al. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science*. 2000; 287(5459):1816–20. Epub 2000/03/10. <https://doi.org/10.1126/science.287.5459.1816> PMID: 10710308.

2. Moriel DG, Bertoldi I, Spagnuolo A, Marchi S, Rosini R, Nesta B, et al. Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(20):9072–7. Epub 2010/05/05. <https://doi.org/10.1073/pnas.0915077107> PMID: 20439758
3. Lauer P, Rinaudo CD, Soriani M, Margarit I, Maione D, Rosini R, et al. Genome analysis reveals pili in Group B *Streptococcus*. *Science*. 2005; 309(5731):105. Epub 2005/07/05. <https://doi.org/10.1126/science.1111563> PMID: 15994549.
4. Nesta B, Spraggon G, Alteri C, Moriel DG, Rosini R, Veggi D, et al. FdeC, a novel broadly conserved *Escherichia coli* adhesin eliciting protection against urinary tract infections. *mBio*. 2012; 3(2). Epub 2012/04/13. <https://doi.org/10.1128/mBio.00010-12> PMID: 22496310
5. McCarthy AJ, Lindsay JA. Genetic variation in *Staphylococcus aureus* surface and immune evasion genes is lineage associated: implications for vaccine design and host-pathogen interactions. *BMC Microbiol*. 2010; 10:173. Epub 2010/06/17. <https://doi.org/10.1186/1471-2180-10-173> PMID: 20550675
6. He M, Sebahia M, Lawley TD, Stabler RA, Dawson LF, Martin MJ, et al. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(16):7527–32. Epub 2010/04/07. <https://doi.org/10.1073/pnas.0914322107> PMID: 20368420
7. Liu X, Afrane M, Clemmer DE, Zhong G, Nelson DE. Identification of *Chlamydia trachomatis* outer membrane complex proteins by differential proteomics. *Journal of bacteriology*. 2010; 192(11):2852–60. Epub 2010/03/30. <https://doi.org/10.1128/JB.01628-09> PMID: 20348250
8. Delany I, Rappuoli R, Seib KL. Vaccines, reverse vaccinology, and bacterial pathogenesis. *Cold Spring Harb Perspect Med*. 2013; 3(5):a012476. Epub 2013/05/03. <https://doi.org/10.1101/cshperspect.a012476> PMID: 23637311
9. Rogovskyy AS, Caoili SEC, Ionov Y, Piontkivska H, Skums P, Tsyvina V, et al. Delineating surface epitopes of Lyme disease pathogen targeted by highly protective antibodies of New Zealand White rabbits. *Infection and immunity*. 2019; 87(8). Epub 2019/05/16. <https://doi.org/10.1128/IAI.00246-19> PMID: 31085705
10. Batool M, Caoili SEC, Dangott LJ, Gerasimov E, Ionov Y, Piontkivska H, et al. Identification of surface epitopes associated with protection against highly immune-evasive VlsE-expressing Lyme disease spirochetes. *Infect Immun*. 2018. Epub 2018/06/06. <https://doi.org/10.1128/IAI.00182-18> PMID: 29866906.
11. Hecker M, Fitzner B, Wendt M, Lorenz P, Flechtner K, Steinbeck F, et al. High-density peptide microarray analysis of IgG autoantibody reactivities in serum and cerebrospinal fluid of multiple sclerosis patients. *Molecular & cellular proteomics: MCP*. 2016; 15(4):1360–80. Epub 2016/02/03. <https://doi.org/10.1074/mcp.M115.051664> PMID: 26831522
12. Chandra A, Latov N, Wormser GP, Marques AR, Alaedini A. Epitope mapping of antibodies to VlsE protein of *Borrelia burgdorferi* in post-Lyme disease syndrome. *Clinical immunology*. 2011; 141(1):103–10. <https://doi.org/10.1016/j.clim.2011.06.005> PMID: 21778118
13. Tapia VE, Ay B, Volkmer R. Exploring and profiling protein function with peptide arrays. *Methods in molecular biology*. 2009; 570:3–17. [https://doi.org/10.1007/978-1-60327-394-7\\_1](https://doi.org/10.1007/978-1-60327-394-7_1) PMID: 19649587.
14. Stafford P, Halperin R, Legutki JB, Magee DM, Galgiani J, Johnston SA. Physical characterization of the "immunosignaturing effect". *Mol Cell Proteomics*. 2012; 11(4):M111 011593. <https://doi.org/10.1074/mcp.M111.011593> PMID: 22261726
15. Stafford P, Cichacz Z, Woodbury NW, Johnston SA. Immunosignature system for diagnosis of cancer. *Proc Natl Acad Sci U S A*. 2014; 111(30):E3072–80. <https://doi.org/10.1073/pnas.1409432111> PMID: 25024171
16. Hughes AK, Cichacz Z, Scheck A, Coons SW, Johnston SA, Stafford P. Immunosignaturing can detect products from molecular markers in brain cancer. *PLoS One*. 2012; 7(7):e40201. <https://doi.org/10.1371/journal.pone.0040201> PMID: 22815729
17. Sykes KF, Legutki JB, Stafford P. Immunosignaturing: a critical review. *Trends Biotechnol*. 2013; 31(1):45–51. <https://doi.org/10.1016/j.tibtech.2012.10.012> PMID: 23219199.
18. Scott JK, Smith GP. Searching for peptide ligands with an epitope library. *Science*. 1990; 249(4967):386–90. Epub 1990/07/27. <https://doi.org/10.1126/science.1696028> PMID: 1696028.
19. Folgori A, Tafi R, Meola A, Felici F, Galfre G, Cortese R, et al. A general strategy to identify mimotopes of pathological antigens using only random peptide libraries and human sera. *EMBO J*. 1994; 13(9):2236–43. Epub 1994/05/01. PMID: 7514533
20. Cortese R, Monaci P, Luzzago A, Santini C, Bartoli F, Cortese I, et al. Selection of biologically active peptides by phage display of random peptide libraries. *Curr Opin Biotechnol*. 1996; 7(6):616–21. Epub 1996/12/01. [https://doi.org/10.1016/s0958-1669\(96\)80072-3](https://doi.org/10.1016/s0958-1669(96)80072-3) PMID: 8939640.

21. Germaschewski V, Murray K. Identification of polyclonal serum specificities with phage-display libraries. *J Virol Methods*. 1996; 58(1–2):21–32. Epub 1996/04/26. [https://doi.org/10.1016/0166-0934\(95\)01980-4](https://doi.org/10.1016/0166-0934(95)01980-4) PMID: 8783147.
22. Scala G, Chen X, Liu W, Telles JN, Cohen OJ, Vaccarezza M, et al. Selection of HIV-specific immunogenic epitopes by screening random peptide libraries with HIV-1-positive sera. *Journal of immunology*. 1999; 162(10):6155–61. Epub 1999/05/07. PMID: 10229859.
23. Mintz PJ, Kim J, Do KA, Wang X, Zinner RG, Cristofanilli M, et al. Fingerprinting the circulating repertoire of antibodies from cancer patients. *Nat Biotechnol*. 2003; 21(1):57–63. Epub 2002/12/24. <https://doi.org/10.1038/nbt774> PMID: 12496764.
24. Zeder-Lutz G, Hoebeke J, Van Regenmortel MH. Differential recognition of epitopes present on monomeric and oligomeric forms of gp160 glycoprotein of human immunodeficiency virus type 1 by human monoclonal antibodies. *Eur J Biochem*. 2001; 268(10):2856–66. Epub 2001/05/19. <https://doi.org/10.1046/j.1432-1327.2001.02167.x> PMID: 11358501.
25. Fukuda MN. Peptide-displaying phage technology in glycobiology. *Glycobiology*. 2012; 22(3):318–25. Epub 2011/09/21. <https://doi.org/10.1093/glycob/cwr140> PMID: 21930649
26. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016; 17(6):333–51. <https://doi.org/10.1038/nrg.2016.49> PMID: 27184599.
27. Matochko WL, Derda R. Next-generation sequencing of phage-displayed peptide libraries. *Methods Mol Biol*. 2015; 1248:249–66. [https://doi.org/10.1007/978-1-4939-2020-4\\_17](https://doi.org/10.1007/978-1-4939-2020-4_17) PMID: 25616338.
28. Ryvkin A, Ashkenazy H, Smelyanski L, Kaplan G, Penn O, Weiss-Ottolenghi Y, et al. Deep panning: steps towards probing the IgOme. *PloS one*. 2012; 7(8):e41469. <https://doi.org/10.1371/journal.pone.0041469> PMID: 22870226
29. Liu X, Hu Q, Liu S, Tallo LJ, Sadzewicz L, Schettine CA, et al. Serum antibody repertoire profiling using *in silico* antigen screen. *PloS one*. 2013; 8(6):e67181. <https://doi.org/10.1371/journal.pone.0067181> PMID: 23826227
30. Christiansen A, Kringelum JV, Hansen CS, Bogh KL, Sullivan E, Patel J, et al. High-throughput sequencing enhanced phage display enables the identification of patient-specific epitope motifs in serum. *Sci Rep*. 2015; 5:12913. <https://doi.org/10.1038/srep12913> PMID: 26246327
31. Rogovskyy AS, Gillis DC, Ionov Y, Gerasimov E, Zelikovsky A. Antibody response to Lyme disease spirochetes in the context of VlsE-mediated immune evasion. *Infection and immunity*. 2017; 85(1). <https://doi.org/10.1128/IAI.00890-16> PMID: 27799330.
32. Batool M, Hillhouse AE, Ionov Y, Kochan KJ, Mohebbi F, Stoica G, et al. New Zealand White rabbits effectively clear *Borrelia burgdorferi* B31 despite the bacteria's functional *vlsE* antigenic variation system. *Infection and immunity*. 2019. Epub 2019/04/17. <https://doi.org/10.1128/IAI.00164-19> PMID: 30988058.
33. Stanek G, Fingerle V, Hunfeld KP, Jaulhac B, Kaiser R, Krause A, et al. Lyme borreliosis: Clinical case definitions for diagnosis and management in Europe. *Clin Microbiol Infect*. 2011; 17(1):69–79. <https://doi.org/10.1111/j.1469-0691.2010.03175.x> PMID: 20132258
34. van den Wijngaard CC, Hofhuis A, Simoes M, Rood E, van Pelt W, Zeller H, et al. Surveillance perspective on Lyme borreliosis across the European Union and European Economic Area. *Euro Surveill*. 2017; 22(27). <https://doi.org/10.2807/1560-7917.ES.2017.22.27.30569> PMID: 28703098.
35. Lindgren E, Jaenson TGT. Lyme borreliosis in Europe: influences of climate and climate change, epidemiology and adaptation measures. In: Menne B, Ebi KL, editors. *Climate change and adaptation strategies for human health*. Darmstadt, Germany: Steinkopff; 2006. p. 157–88.
36. Hubalek Z. Epidemiology of Lyme borreliosis. *Curr Probl Dermatol*. 2009; 37:31–50. <https://doi.org/10.1159/000213069> PMID: 19367096.
37. Hinckley AF, Connally NP, Meek JL, Johnson BJ, Kemperman MM, Feldman KA, et al. Lyme disease testing by large commercial laboratories in the United States. *Clin Infect Dis*. 2014; 59(5):676–81. <https://doi.org/10.1093/cid/ciu397> PMID: 24879782
38. Littman MP, Gerber B, Goldstein RE, Labato MA, Lappin MR, Moore GE. ACVIM consensus update on Lyme borreliosis in dogs and cats. *J Vet Intern Med*. 2018; 32(3):887–903. Epub 2018/03/23. <https://doi.org/10.1111/jvim.15085> PMID: 29566442
39. Brownstein JS, Holford TR, Fish D. Effect of climate change on Lyme disease risk in North America. *Ecohealth*. 2005; 2(1):38–46. Epub 2008/11/15. <https://doi.org/10.1007/s10393-004-0139-x> PMID: 19008966
40. Ostfeld RS, Brunner JL. Climate change and *Ixodes* tick-borne diseases of humans. *Philos Trans R Soc Lond B Biol Sci*. 2015;370(1665). Epub 2015/02/18. <https://doi.org/10.1098/rstb.2014.0051> PMID: 25688022

41. Rizzoli A, Hauffe HC, Carpi G, Vourc'h GI, Neteler M, Rosa R. Lyme borreliosis in Europe. *Eurosurveillance*. 2011; 16(27):2–9.
42. Peavey CA, Lane RS. Transmission of *Borrelia burgdorferi* by *Ixodes pacificus* nymphs and reservoir competence of deer mice (*Peromyscus maniculatus*) infected by tick-bite. *J Parasitol*. 1995; 81(2):175–8. Epub 1995/04/01. PMID: 7707191.
43. Rand PW, Lacombe EH, Smith RP Jr., Rich SM, Kilpatrick CW, Dragoni CA, et al. Competence of *Peromyscus maniculatus* (Rodentia: Cricetidae) as a reservoir host for *Borrelia burgdorferi* (Spirochaetales: Spirochaetaceae) in the wild. *Journal of medical entomology*. 1993; 30(3):614–8. Epub 1993/05/01. <https://doi.org/10.1093/jmedent/30.3.614> PMID: 8510121.
44. Radolf JD, Caimano MJ, Stevenson B, Hu LT. Of ticks, mice and men: understanding the dual-host life-style of Lyme disease spirochaetes. *Nat Rev Microbiol*. 2012; 10(2):87–99. Epub 2012/01/11. <https://doi.org/10.1038/nrmicro2714> PMID: 22230951
45. Anderson JF, Johnson RC, Magnarelli LA, Hyde FW. Culturing *Borrelia burgdorferi* from spleen and kidney tissues of wild-caught white-footed mice, *Peromyscus leucopus*. *Zentralbl Bakteriol Mikrobiol Hyg A*. 1986; 263(1–2):34–9. Epub 1986/12/01. [https://doi.org/10.1016/s0176-6724\(86\)80099-2](https://doi.org/10.1016/s0176-6724(86)80099-2) PMID: 3577490.
46. Anderson JF, Johnson RC, Magnarelli LA. Seasonal prevalence of *Borrelia burgdorferi* in natural populations of white-footed mice, *Peromyscus leucopus*. *Journal of clinical microbiology*. 1987; 25(8):1564–6. Epub 1987/08/01. PMID: 3624451
47. Schwan TG, Kime KK, Schrupf ME, Coe JE, Simpson WJ. Antibody response in white-footed mice (*Peromyscus leucopus*) experimentally infected with the Lyme disease spirochete (*Borrelia burgdorferi*). *Infection and immunity*. 1989; 57(11):3445–51. Epub 1989/11/01. PMID: 2807530
48. Hudson BJ, Stewart M, Lennox VA, Fukunaga M, Yabuki M, Macorison H, et al. Culture-positive Lyme borreliosis. *Med J Aust*. 1998; 168(10):500–2. Epub 1998/06/19. PMID: 9631675.
49. Oksi J, Marjamaki M, Nikoskelainen J, Viljanen MK. *Borrelia burgdorferi* detected by culture and PCR in clinical relapse of disseminated Lyme borreliosis. *Ann Med*. 1999; 31(3):225–32. Epub 1999/08/12. <https://doi.org/10.3109/07853899909115982> PMID: 10442678.
50. Kash N, Fink-Puches R, Cerroni L. Cutaneous manifestations of B-cell chronic lymphocytic leukemia associated with *Borrelia burgdorferi* infection showing a marginal zone B-cell lymphoma-like infiltrate. *Am J Dermatopathol*. 2011; 33(7):712–5. Epub 2011/09/29. <https://doi.org/10.1097/DAD.0b013e3181fc576f> PMID: 21946761.
51. Middelveen MJ, Sapi E, Burke J, Filush KR, Franco A, Fesler MC, et al. Persistent *Borrelia* infection in patients with ongoing symptoms of Lyme disease. *Healthcare (Basel)*. 2018; 6(2). Epub 2018/04/18. <https://doi.org/10.3390/healthcare6020033> PMID: 29662016
52. Logigian EL, Kaplan RF, Steere AC. Chronic neurologic manifestations of Lyme disease. *The New England journal of medicine*. 1990; 323(21):1438–44. Epub 1990/11/22. <https://doi.org/10.1056/NEJM19901123232102> PMID: 2172819.
53. Stanek G, Klein J, Bittner R, Glogar D. Isolation of *Borrelia burgdorferi* from the myocardium of a patient with longstanding cardiomyopathy. *The New England journal of medicine*. 1990; 322(4):249–52. Epub 1990/01/25. <https://doi.org/10.1056/NEJM199001253220407> PMID: 2294450.
54. Rogovskyy AS, Bankhead T. Variable VlsE is critical for host reinfection by the Lyme disease spirochete. *PloS one*. 2013; 8(4):e61226. <https://doi.org/10.1371/journal.pone.0061226> PMID: 23593438
55. Vaz A, Glickstein L, Field JA, McHugh G, Sikand VK, Damle N, et al. Cellular and humoral immune responses to *Borrelia burgdorferi* antigens in patients with culture-positive early Lyme disease. *Infection and immunity*. 2001; 69(12):7437–44. Epub 2001/11/14. <https://doi.org/10.1128/IAI.69.12.7437-7444.2001> PMID: 11705918
56. LaRocca TJ, Benach JL. The important and diverse roles of antibodies in the host response to *Borrelia* infections. *Curr Top Microbiol Immunol*. 2008; 319:63–103. [https://doi.org/10.1007/978-3-540-73900-5\\_4](https://doi.org/10.1007/978-3-540-73900-5_4) PMID: 18080415.
57. Xu Y, Bruno JF, Luft BJ. Profiling the humoral immune response to *Borrelia burgdorferi* infection with protein microarrays. *Microbial pathogenesis*. 2008; 45(5–6):403–7. <https://doi.org/10.1016/j.micpath.2008.09.006> PMID: 18976702.
58. Lawrenz MB, Hardham JM, Owens RT, Nowakowski J, Steere AC, Wormser GP, et al. Human antibody responses to VlsE antigenic variation protein of *Borrelia burgdorferi*. *Journal of clinical microbiology*. 1999; 37(12):3997–4004. Epub 1999/11/24. PMID: 10565921
59. McDowell JV, Sung SY, Hu LT, Marconi RT. Evidence that the variable regions of the central domain of VlsE are antigenic during infection with Lyme disease spirochetes. *Infection and immunity*. 2002; 70(8):4196–203. <https://doi.org/10.1128/IAI.70.8.4196-4203.2002> PMID: 12117928.

60. Fikrig E, Bockenstedt LK, Barthold SW, Chen M, Tao H, Ali-Salaam P, et al. Sera from patients with chronic Lyme disease protect mice from Lyme borreliosis. *The Journal of infectious diseases*. 1994; 169(3):568–74. Epub 1994/03/01. <https://doi.org/10.1093/infdis/169.3.568> PMID: 8158028.
61. Zhang JR, Hardham JM, Barbour AG, Norris SJ. Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. *Cell*. 1997; 89(2):275–85. [https://doi.org/10.1016/s0092-8674\(00\)80206-8](https://doi.org/10.1016/s0092-8674(00)80206-8) PMID: 9108482
62. Coutte L, Botkin DJ, Gao L, Norris SJ. Detailed analysis of sequence changes occurring during *visE* antigenic variation in the mouse model of *Borrelia burgdorferi* infection. *PLoS pathogens*. 2009; 5(2): e1000293. Epub 2009/02/14. <https://doi.org/10.1371/journal.ppat.1000293> PMID: 19214205
63. Purser JE, Norris SJ. Correlation between plasmid content and infectivity in *Borrelia burgdorferi*. *Proceedings of the National Academy of Sciences of the United States of America*. 2000; 97(25):13865–70. Epub 2000/12/06. <https://doi.org/10.1073/pnas.97.25.13865> PMID: 11106398
64. Labandeira-Rey M, Skare JT. Decreased infectivity in *Borrelia burgdorferi* strain B31 is associated with loss of linear plasmid 25 or 28–1. *Infection and immunity*. 2001; 69(1):446–55. <https://doi.org/10.1128/IAI.69.1.446-455.2001> PMID: 11119536
65. Labandeira-Rey M, Seshu J, Skare JT. The absence of linear plasmid 25 or 28–1 of *Borrelia burgdorferi* dramatically alters the kinetics of experimental infection via distinct mechanisms. *Infection and immunity*. 2003; 71(8):4608–13. Epub 2003/07/23. <https://doi.org/10.1128/IAI.71.8.4608-4613.2003> PMID: 12874340
66. Iyer R, Kalu O, Purser J, Norris S, Stevenson B, Schwartz I. Linear and circular plasmid content in *Borrelia burgdorferi* clinical isolates. *Infection and immunity*. 2003; 71(7):3699–706. <https://doi.org/10.1128/IAI.71.7.3699-3706.2003> PMID: 12819050.
67. Lawrenz MB, Wooten RM, Norris SJ. Effects of *visE* complementation on the infectivity of *Borrelia burgdorferi* lacking the linear plasmid lp28-1. *Infection and immunity*. 2004; 72(11):6577–85. <https://doi.org/10.1128/IAI.72.11.6577-6585.2004> PMID: 15501789.
68. Bankhead T, Chaconas G. The role of *VisE* antigenic variation in the Lyme disease spirochete: persistence through a mechanism that differs from other pathogens. *Molecular microbiology*. 2007; 65(6):1547–58. <https://doi.org/10.1111/j.1365-2958.2007.05895.x> PMID: 17714442.
69. Rogovskyy AS, Casselli T, Tourand Y, Jones CR, Owen JP, Mason KL, et al. Evaluation of the importance of *VisE* antigenic variation for the enzootic cycle of *Borrelia burgdorferi*. *PloS one*. 2015; 10(4): e0124268. <https://doi.org/10.1371/journal.pone.0124268> PMID: 25893989
70. Barthold SW, de Souza MS, Janotka JL, Smith AL, Persing DH. Chronic Lyme borreliosis in the laboratory mouse. *Am J Pathol*. 1993; 143(3):959–71. Epub 1993/09/01. PMID: 8362988
71. Johnson RC, Marek N, Kodner C. Infection of Syrian hamsters with Lyme disease spirochetes. *J Clin Microbiol*. 1984; 20(6):1099–101. Epub 1984/12/01. PMID: 6520220
72. Goodman JL, Jurkovich P, Kodner C, Johnson RC. Persistent cardiac and urinary tract infections with *Borrelia burgdorferi* in experimentally infected Syrian hamsters. *Journal of clinical microbiology*. 1991; 29(5):894–6. Epub 1991/05/01. PMID: 2056054
73. Burgess EC. Experimental inoculation of dogs with *Borrelia burgdorferi*. *Zentralbl Bakteriol Mikrobiol Hyg A*. 1986; 263(1–2):49–54. Epub 1986/12/01. [https://doi.org/10.1016/s0176-6724\(86\)80102-x](https://doi.org/10.1016/s0176-6724(86)80102-x) PMID: 3554844.
74. Greene RT, Levine JF, Breitschwerdt EB, Walker RL, Berkhoff HA, Cullen J, et al. Clinical and serologic evaluations of induced *Borrelia burgdorferi* infection in dogs. *Am J Vet Res*. 1988; 49(6):752–7. Epub 1988/06/01. PMID: 3041881.
75. Straubinger RK, Summers BA, Chang YF, Appel MJ. Persistence of *Borrelia burgdorferi* in experimentally infected dogs after antibiotic treatment. *J Clin Microbiol*. 1997; 35(1):111–6. Epub 1997/01/01. PMID: 8968890
76. Straubinger RK. PCR-Based quantification of *Borrelia burgdorferi* organisms in canine tissues over a 500-Day postinfection period. *J Clin Microbiol*. 2000; 38(6):2191–9. Epub 2000/06/02. PMID: 10834975
77. Appel MJ, Allan S, Jacobson RH, Lauderdale TL, Chang YF, Shin SJ, et al. Experimental Lyme disease in dogs produces arthritis and persistent infection. *J Infect Dis*. 1993; 167(3):651–64. Epub 1993/03/01. <https://doi.org/10.1093/infdis/167.3.651> PMID: 8440936.
78. Preac Mursic V, Patsouris E, Wilske B, Reinhardt S, Gross B, Mehraein P. Persistence of *Borrelia burgdorferi* and histopathological alterations in experimentally infected animals. A comparison with histopathological findings in human Lyme disease. *Infection*. 1990; 18(6):332–41. Epub 1990/11/01. <https://doi.org/10.1007/bf01646399> PMID: 2076905.
79. Sonnesyn SW, Manivel JC, Johnson RC, Goodman JL. A guinea pig model for Lyme disease. *Infect Immun*. 1993; 61(11):4777–84. Epub 1993/11/01. PMID: 8406878

80. Philipp MT, Aydintug MK, Bohm RP Jr., Cogswell FB, Dennis VA, Lanners HN, et al. Early and early disseminated phases of Lyme disease in the rhesus monkey: a model for infection in humans. *Infection and immunity*. 1993; 61(7):3047–59. PMID: [8514412](#)
81. Roberts ED, Bohm RP Jr., Cogswell FB, Lanners HN, Lowrie RC Jr., Povinelli L, et al. Chronic lyme disease in the rhesus monkey. *Laboratory investigation; a journal of technical methods and pathology*. 1995; 72(2):146–60. PMID: [7853849](#).
82. Barthold SW, Beck DS, Hansen GM, Terwilliger GA, Moody KD. Lyme borreliosis in selected strains and ages of laboratory mice. *The Journal of infectious diseases*. 1990; 162(1):133–8. <https://doi.org/10.1093/infdis/162.1.133> PMID: [2141344](#).
83. Hodzic E, Feng S, Holden K, Freet KJ, Barthold SW. Persistence of *Borrelia burgdorferi* following antibiotic treatment in mice. *Antimicrob Agents Chemother*. 2008; 52(5):1728–36. <https://doi.org/10.1128/AAC.01050-07> PMID: [18316520](#)
84. Chang YF, Ku YW, Chang CF, Chang CD, McDonough SP, Divers T, et al. Antibiotic treatment of experimentally *Borrelia burgdorferi*-infected ponies. *Vet Microbiol*. 2005; 107(3–4):285–94. Epub 2005/05/03. <https://doi.org/10.1016/j.vetmic.2005.02.006> PMID: [15863289](#).
85. Chang YF, Novosol V, McDonough SP, Chang CF, Jacobson RH, Divers T, et al. Experimental infection of ponies with *Borrelia burgdorferi* by exposure to Ixodid ticks. *Vet Pathol*. 2000; 37(1):68–76. Epub 2000/01/22. <https://doi.org/10.1354/vp.37-1-68> PMID: [10643983](#).
86. Barthold SW, Moody KD, Terwilliger GA, Duray PH, Jacoby RO, Steere AC. Experimental Lyme arthritis in rats infected with *Borrelia burgdorferi*. *The Journal of infectious diseases*. 1988; 157(4):842–6. <https://doi.org/10.1093/infdis/157.4.842> PMID: [3258003](#).
87. Foley DM, Gayek RJ, Skare JT, Wagar EA, Champion CI, Blanco DR, et al. Rabbit model of Lyme borreliosis: erythema migrans, infection-derived immunity, and identification of *Borrelia burgdorferi* proteins associated with virulence and protective immunity. *J Clin Invest*. 1995; 96(2):965–75. <https://doi.org/10.1172/JCI118144> PMID: [7635989](#)
88. Embers ME, Liang FT, Howell JK, Jacobs MB, Purcell JE, Norris SJ, et al. Antigenicity and recombination of VlsE, the antigenic variation protein of *Borrelia burgdorferi*, in rabbits, a host putatively resistant to long-term infection with this spirochete. *FEMS immunology and medical microbiology*. 2007; 50(3):421–9. Epub 2007/06/29. <https://doi.org/10.1111/j.1574-695X.2007.00276.x> PMID: [17596185](#).
89. Liang FT, Brown EL, Wang T, Iozzo RV, Fikrig E. Protective niche for *Borrelia burgdorferi* to evade humoral immunity. *Am J Pathol*. 2004; 165(3):977–85. Epub 2004/08/28. [https://doi.org/10.1016/S0002-9440\(10\)63359-7](https://doi.org/10.1016/S0002-9440(10)63359-7) PMID: [15331421](#)
90. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*. 2009; 37(Web Server issue):W202–8. Epub 2009/05/22. <https://doi.org/10.1093/nar/gkp335> PMID: [19458158](#)
91. Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic acids research*. 2015; 43(W1):W566–70. Epub 2015/05/15. <https://doi.org/10.1093/nar/gkv468> PMID: [25969447](#)
92. Sette A, Rappuoli R. Reverse vaccinology: developing vaccines in the era of genomics. *Immunity*. 2010; 33(4):530–41. Epub 2010/10/30. <https://doi.org/10.1016/j.immuni.2010.09.017> PMID: [21029963](#)
93. Rogovskyy AS, Caoili SEC, Ionov Y, Piontkivska H, Skums P, Tsyvina V, et al. Delineating surface epitopes of *Borrelia burgdorferi* targeted by highly protective antibody of the rabbit. *Infection and immunity*. 2019.
94. Wu CH, Liu IJ, Lu RM, Wu HC. Advancement and applications of peptide phage display technology in biomedical science. *J Biomed Sci*. 2016; 23:8. <https://doi.org/10.1186/s12929-016-0223-x> PMID: [26786672](#)
95. Legutki JB, Zhao ZG, Greving M, Woodbury N, Johnston SA, Stafford P. Scalable high-density peptide arrays for comprehensive health monitoring. *Nat Commun*. 2014; 5:4785. <https://doi.org/10.1038/ncomms5785> PMID: [25183057](#).
96. Bastas G, Sompuram SR, Pierce B, Vani K, Bogen SA. Bioinformatic requirements for protein database searching using predicted epitopes from disease-associated antibodies. *Molecular & cellular proteomics: MCP*. 2008; 7(2):247–56. <https://doi.org/10.1074/mcp.M700107-MCP200> PMID: [17897933](#).
97. Ionov Y. A high throughput method for identifying personalized tumor-associated antigens. *Oncotarget*. 2010; 1(2):148–55. <https://doi.org/10.18632/oncotarget.118> PMID: [20711419](#)
98. Paull ML, Johnston T, Ibsen KN, Bozekowski JD, Daugherty PS. A general approach for predicting protein epitopes targeted by antibody repertoires using whole proteomes. *PloS one*. 2019; 14(9): e0217668. Epub 2019/09/07. <https://doi.org/10.1371/journal.pone.0217668> PMID: [31490930](#).
99. Elias AF, Stewart PE, Grimm D, Caimano MJ, Eggers CH, Tilly K, et al. Clonal polymorphism of *Borrelia burgdorferi* strain B31 MI: Implications for mutagenesis in an infectious strain background. *Infection and immunity*. 2002; 70(4):2139–50. <https://doi.org/10.1128/IAI.70.4.2139-2150.2002> PMID: [11895980](#).