



Protein-Based Immunome Wide Association Studies (PIWAS) for the Discovery of Significant Disease-Associated Antigens

Winston A. Haynes, Kathy Kamath, Rebecca Waitz, Patrick S. Daugherty and John C. Shon*

Serimmune, Inc., Santa Barbara, CA, United States

OPEN ACCESS

Edited by:

Deborah K. Dunn-Walters,
University of Surrey,
United Kingdom

Reviewed by:

Phillip Cohen,
Temple University,
United States
Natalia V. Giltiay,
University of Washington,
United States

*Correspondence:

John C. Shon
john.shon@serimmune.com

Specialty section:

This article was submitted to
B Cell Biology,
a section of the journal
Frontiers in Immunology

Received: 02 November 2020

Accepted: 07 April 2021

Published: 27 April 2021

Citation:

Haynes WA, Kamath K, Waitz R,
Daugherty PS and Shon JC (2021)
Protein-Based Immunome Wide
Association Studies (PIWAS) for
the Discovery of Significant
Disease-Associated Antigens.
Front. Immunol. 12:625311
doi: 10.3389/fimmu.2021.625311

Identification of the antigens associated with antibodies is vital to understanding immune responses in the context of infection, autoimmunity, and cancer. Discovering antigens at a proteome scale could enable broader identification of antigens that are responsible for generating an immune response or driving a disease state. Although targeted tests for known antigens can be straightforward, discovering antigens at a proteome scale using protein and peptide arrays is time consuming and expensive. We leverage Serum Epitope Repertoire Analysis (SERA), an assay based on a random bacterial display peptide library coupled with next generation sequencing (NGS), to power the development of Protein-based Immunome Wide Association Study (PIWAS). PIWAS uses proteome-based signals to discover candidate antibody-antigen epitopes that are significantly elevated in a subset of cases compared to controls. After demonstrating statistical power relative to the magnitude and prevalence of effect in synthetic data, we apply PIWAS to systemic lupus erythematosus (SLE, n=31) and observe known autoantigens, Smith and Ribosomal protein P, within the 22 highest scoring candidate protein antigens across the entire human proteome. We validate the magnitude and location of the SLE specific signal against the Smith family of proteins using a cohort of patients who are positive by predicate anti-Sm tests. To test the generalizability of the method in an additional autoimmune disease, we identified and validated autoantigenic signals to SSB, CENPA, and keratin proteins in a cohort of individuals with Sjogren's syndrome (n=91). Collectively, these results suggest that PIWAS provides a powerful new tool to discover disease-associated serological antigens within any known proteome.

Keywords: computational immunology, antigen, immunome, bioinformatics & computational biology, immunology & inflammation

INTRODUCTION

Antibodies present in human specimens serve as the primary analyte and disease biomarker for a broad group of infectious (bacterial, viral, fungal, and parasitic) and autoimmune diseases. As such, hundreds of distinct antibody-detecting immunoassays have been developed to diagnose human disease using blood derived specimens. The development of high-throughput sequencing technologies has enabled sequencing of numerous proteomes from diverse organisms. However, methods for antigen discovery within any given proteome remain relatively low throughput. The serological analysis of expression cDNA libraries (SEREX) method has been applied frequently to identify a variety of antigens, but high-quality cDNA library construction remains technically challenging and time consuming (1–3). Alternatively, entire human and pathogen derived proteomes can be segmented into overlapping peptides, and displayed on phage or solid-phase arrays and probed with serum (4–6). Fully random peptide arrays of up to 300,000 unique sequences have also been used successfully to detect antibodies towards a range of organisms (7–9). Even so, the limited molecular diversity of array based libraries can reduce antibody detection sensitivity and hinder successful mapping of peptide motifs to specific proteome antigens (7). Thus, a general, scalable approach to identify serological antigens within arbitrary proteomes is needed.

In autoimmune diseases and cancers, autoantigen discovery is further complicated by the size of the proteome, heterogeneity of disease, and variability in immune response. Patient genetics, exposures, and microbiomes contribute to this heterogeneity, which in turn yields disparate responses to diverse antigens and epitopes (10, 11). In such cases, the mapping of multiple epitopes to one antigen can increase confidence in a candidate antigen (7, 12). Even for diseases with conserved autoantigens, epitope spreading can lead to a diversified immune response against additional epitopes from the same protein or other proteins from the same tissue (13, 14). In cancer patients, neoepitopes can arise in response to somatic mutations that yield conformational changes or abnormal expression (15, 16).

In complex autoimmune diseases like systemic lupus erythematosus (SLE) and Sjogren's syndrome, autoantibodies play an important role in diagnosis, patient stratification, and pathogenesis. SLE autoantigens include double-stranded DNA, ribonuclear proteins (Smith), C1q, α -actinin, α -enolase, annexin II, annexin AI, and ribosomal protein P (17–19). In particular, anti-Smith antigen antibodies are present in 25–30% of SLE patients (20, 21). The Smith antigen consists of a complex of U-rich RNA U1, U2, U4/U6, and U5, along with core polypeptides B', B, D1, D2, D3, E, F, and G. Not all components of this complex are equally antigenic, and there are multiple epitopes within the complex (22, 23). Prominent Sjogren's syndrome autoantigens include Ro/SSA, Lupus La protein/SSB, salivary gland protein 1 and parotid secretory protein (24, 25). Additional families of antibodies include: anti-nuclear, rheumatoid factor, anti-keratin, anti-centromere, anti-mitochondrial and anti-cyclic citrullinated peptides (24, 25).

One approach for antigen discovery, serum epitope repertoire analysis (SERA), uses bacterial display technology to present random 12mer peptides to serum antibodies (26–28). Peptides expressed on the surface of the bacteria bind to serum antibodies and are separated using protein A/G magnetic beads. Plasmids encoding the peptides are purified and the peptide-encoding regions are amplified and sequenced using next-generation sequencing. For each of these peptides and their kmer subsequences, enrichment can be calculated by comparing the actual number of observations to that expected based on amino acid frequencies (26). Mapping these peptide epitopes to their corresponding protein antigens requires protein structure and/or sequence. Structure-based epitope mapping methods (e.g., 3DEX, MIMOX, MIMOP, Pepitope) are not yet feasible at a proteome scale, due in part to the large number of undetermined structures (29–32). However, since 85% of epitope-paratope interactions in crystal structures have a linear stretch of 5 amino acids, sequence information alone can be sufficient to identify many antigens (33–35). The K-TOPE (Kmer-Tiling of Protein Epitopes) method has demonstrated the ability of tiled 5-mers to identify known epitopes in a variety of infections at proteome scale (36). Here, we present a method, Protein-based Immunome Wide Association Study (PIWAS), which leverages the SERA assay to discover disease relevant antigens within large cohorts and at proteome scale. We evaluate PIWAS with synthetic data to examine the magnitude and prevalence of the effect needed for robust detection. We validate PIWAS using specimens from individuals with SLE and controls, identifying established anti-Smith and anti-Ribosomal protein P autoantibodies. We further validate the anti-Smith epitopes identified in our analysis using specimens positive for anti-Smith autoantibodies by predicate tests.

METHOD

PIWAS Allows Identification of Proteome-Based Signals

To identify candidate serological antigens from arbitrary proteomes, we developed a robust, cohort-based statistical method to analyze peptide sequence data from the SERA assay. SERA uses a large bacterial display random peptide library of 10 billion member 12mers to identify binding to the epitopes recognized by antibodies species in a biospecimen (e.g. serum, plasma, cerebrospinal fluid) (Figure 1A). From a typical specimen, we acquire 1–5 million unique 12mers. We break these 12mers into their constituent kmers, calculate log-enrichments (observed divided by expected counts), and store the results in a BigTable database. To identify disease-specific antigens from these data, PIWAS compares kmer data from case and control cohorts against a proteome of interest (Figure 1B). For each protein and specimen dataset, we calculate tiled kmer enrichments (normalized to the controls as a background) and smooth across a sliding window. For each protein, we leverage statistics such as the outlier sum and Mann-Whitney test to compare the case and

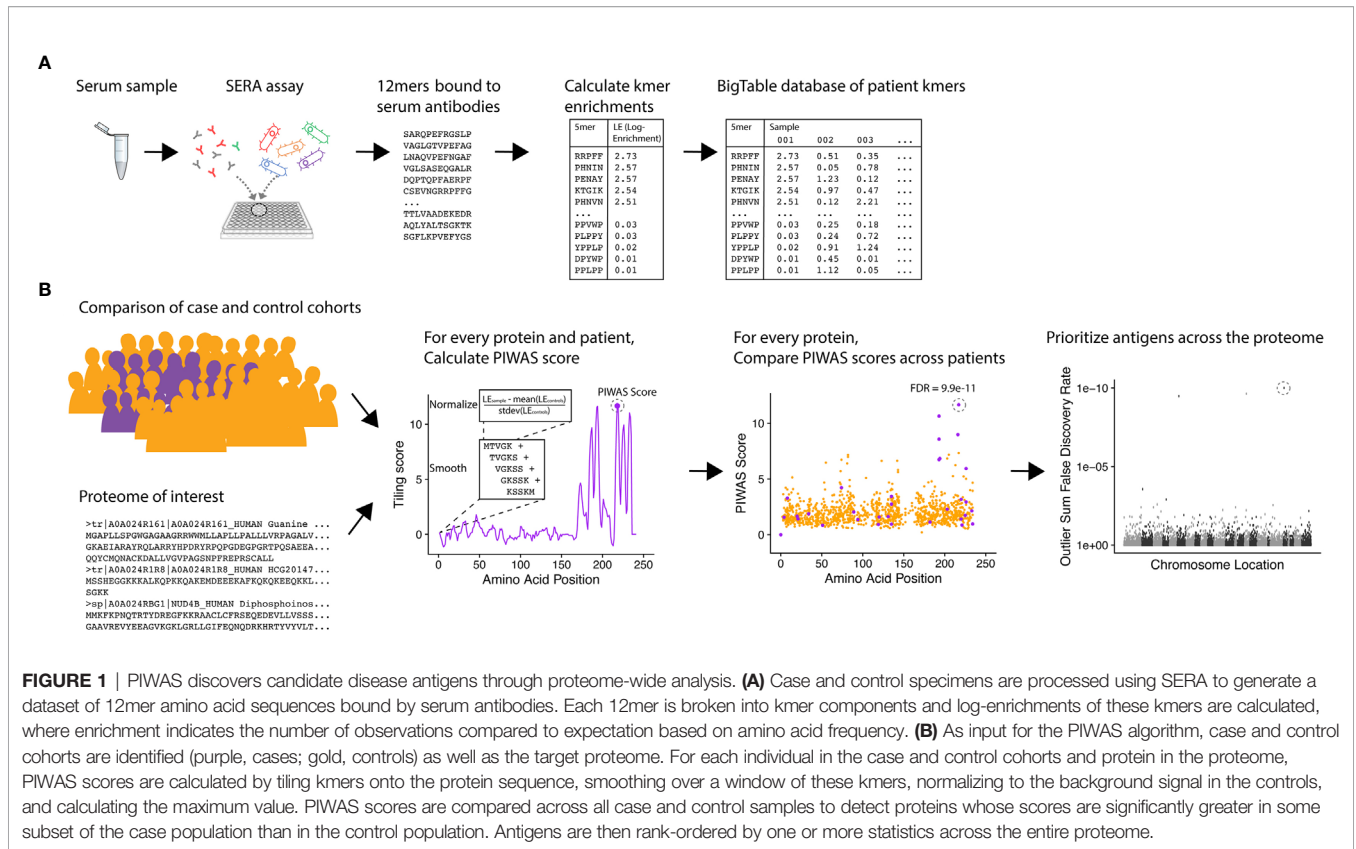


FIGURE 1 | PIWAS discovers candidate disease antigens through proteome-wide analysis. **(A)** Case and control specimens are processed using SERA to generate a dataset of 12mer amino acid sequences bound by serum antibodies. Each 12mer is broken into kmer components and log-enrichments of these kmers are calculated, where enrichment indicates the number of observations compared to expectation based on amino acid frequency. **(B)** As input for the PIWAS algorithm, case and control cohorts are identified (purple, cases; gold, controls) as well as the target proteome. For each individual in the case and control cohorts and protein in the proteome, PIWAS scores are calculated by tiling kmers onto the protein sequence, smoothing over a window of these kmers, normalizing to the background signal in the controls, and calculating the maximum value. PIWAS scores are compared across all case and control samples to detect proteins whose scores are significantly greater in some subset of the case population than in the control population. Antigens are then rank-ordered by one or more statistics across the entire proteome.

control populations. At a proteome scale, we prioritize candidate antigens based on these statistics (see details below).

PIWAS Calculation

We define case (T), and control (U), cohorts of samples and begin with 12mer amino acid sequences for each sample generated by SERA (minimum of $1e6$ total unique sequences per sample).

Enrichment Calculation

We decompose each 12 mer from SERA into constituent kmers (where $k=5$ and $k=6$ consecutive amino acids). For every $kmer$ in each sample (S), we calculate enrichment as:

$$E_s(kmer) = n_s(kmer) / e_s(kmer)$$

where $n(kmer)$ is the number of unique 12mers containing a particular $kmer$ and $e_s(kmer)$ is the expected number of $kmer$ reads for the sample, defined as:

$$e_s(kmer) = N_s (L_{seq} - k + 1) \prod_{i=1}^k p_i$$

where N_s is the number of 12mer reads generated for S , L_{seq} is the length of the amino acid reads (12), k is the kmer length, and p_i is the amino acid proportion for the i th amino acid in $kmer$ in all 12mers from S .

Number of Standard Deviation Normalization

For every kmer, we normalize enrichment values to a control population. We define the control enrichment values as:

$$C = \{E_v(kmer) : v \in W\}$$

where W is the control cohort (U).

The normalized enrichment is calculated as:

$$F_s(kmer) = \frac{E_s(kmer) - \mu(C)}{\sigma(C)}$$

where $\mu(C)$ is the mean of C and $\sigma(C)$ is the standard deviation of C .

PIWAS Score Calculation

For each protein p and sample s , we calculate a PIWAS score $P(s, p)$, defined as:

$$P(s, p) = \max_{1 \leq i \leq len(p)} \sum_{k=5}^6 \sum_{j=i}^{\min(i+w, len(p)-k)} G_s(kmer(j, k, p))$$

where w is the width of the smoothing window, $len(p)$ is the length of protein p , $kmer(j, k, p)$ is the kmer of length k at location j in protein p , and G_s is either E_s or F_s . Similarly, we record the location of this maximum statistics value, $P_{loc}(s, p)$, as:

$$P_{loc}(s, p) = \operatorname{argmax}_{1 \leq i \leq \operatorname{len}(p)} \sum_{k=5}^{\min(i+w, \operatorname{len}(p)-k)} \sum_{j=i} G_S(\operatorname{kmer}(j, k, p))$$

$$z_{O(p)} = \frac{O(p) - \mu_{O^0(p)}}{\sigma_{O^0(p)}}$$

Cohort Comparison Statistics

For each protein p , we define our case enrichments as:

$$A(p) = \{P(t, p) : t \in T\}$$

Similarly, we define our control enrichments as:

$$B(p) = \{P(u, p) : u \in U\}$$

We use several statistical tests to compare $A(p)$ and $B(p)$, including traditional tests like the Mann-Whitney U and Kolmogorov-Smirnov. We calculate effect size as the Hedges' g statistic.

We calculate the Outlier Sum, which we define as $O(p)$, statistic defined in Tibshirani and Hastie (37). We perform 1,000 random permutations of the samples in $A(p)$ and $B(p)$ and calculate the Outlier Sum to calculate $O^0(p)$, the null distribution of the Outlier Sum for protein p . We calculate the z-score as:

Since the Outlier Sum is a sum of i.i.d. variables, we can apply the Central Limit Theorem and calculate a p-value for $z_{O(p)}$ using the normal distribution.

We define the sets of case and control locations as:

$$A_{loc}(p) = \{P_{loc}(t, p) : t \in T\}$$

$$B_{loc}(p) = \{P_{loc}(u, p) : u \in U\}$$

We perform a Kolmogorov-Smirnov test comparing $A_{loc}(p)$ $B_{loc}(p)$ and to identify proteins with locational conservation of epitopes.

RESULTS

Kmer Enrichment in Samples with Serum Compared to Enrichment in a Random Library

We first compared SERA library sequence composition before and after library selection with serum from healthy controls and SLE patients (Figures 2A, B). Both the control and SLE serum yielded

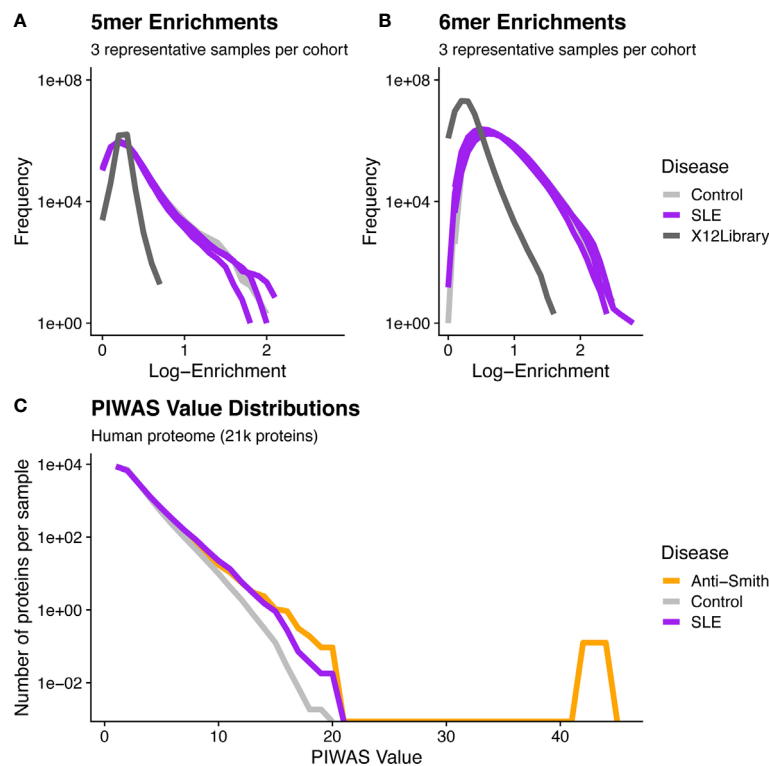


FIGURE 2 | Distributional differences in kmer enrichments and PIWAS values between the unselected library and after selecting with SLE and control specimens. 5mer (A) and 6mer (B) Kmer frequency (y-axis) vs. Log-enrichment score (x-axis) for 6 subjects and the naïve library demonstrates species with large enrichments are found exclusively in those SERA assays incubated with serum. All 5mers or 6mers from three representative samples per cohort are evaluated for enrichment. Dark-gray lines = naïve 12-mer peptide library, purple lines = SLE cohort, gray lines = control cohort. (C) A comparison of PIWAS values (x-axis) vs. the number of proteins per sample with the corresponding PIWAS value (y-axis) reveals differences in both the range and distribution of PIWAS values between SLE and control samples. Distributions are based on 31 SLE cases and 1,157 controls. Purple = SLE cohort, gray = control cohort, orange = anti-Smith cohort.

larger enrichments for both 5mers and 6mers compared to the unselected library. The enrichment of 5mers and 6mers in samples incubated with serum demonstrates the effects of antibody selection on the peptide library composition. We also compared the distribution of PIWAS values when 5mers were mapped to the human proteome. Interestingly, both SLE and anti-Smith cohorts yielded PIWAS value distributions with longer tails when analyzed against the entire human proteome when compared to those of healthy controls (**Figures 2C**). These findings confirm the general basis for using 5mers and 6mers for identifying both enriched signal in serum relative to a random library and enriched autoantigen signal using PIWAS in an example disease population relative to healthy controls.

PIWAS Power Simulations

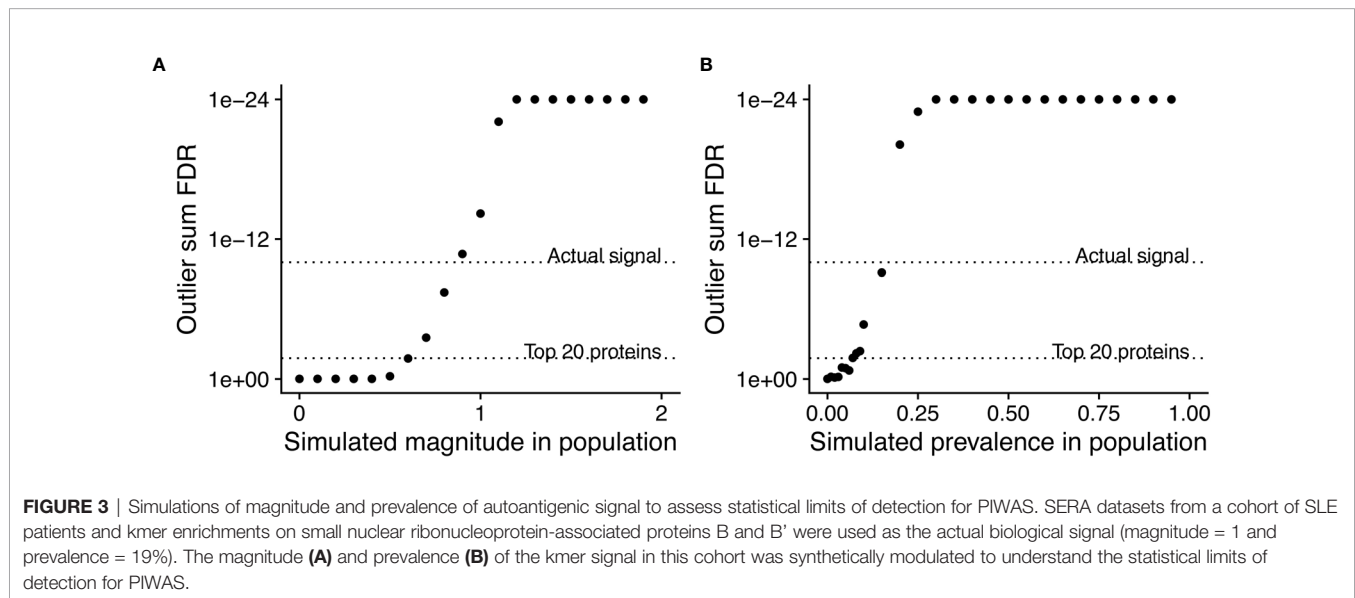
In order to assess the statistical power of PIWAS to detect enriched antigens in a cohort, we performed computational experiments where we adjusted the magnitude and prevalence of known autoantigenic signal against Sm antigens (specifically small nuclear ribonucleoprotein-associated proteins B and B') in a cohort of SLE patients. Unsurprisingly, as the magnitude of the effect increases, so does the significance of the antigenic signal (**Figure 3A**). At an effect of only 60% of the SERA signal obtained with true SLE biospecimens, Sm antigens are significant at a false discovery rate (FDR)=0.017 using the outlier sum FDR, still ranking within the top 20 proteins. Similarly, as the prevalence of the anti-Sm signal increases in the case population, so too does the significance of the outlier sum p-value (**Figure 3B**). At a prevalence of 7% (less than half of the actual biological prevalence in this cohort), anti-Sm is significant at FDR= 0.015 and remains within the top 20 scoring proteins. These results indicate an ability to detect signals well below the prevalence of many established autoantigens.

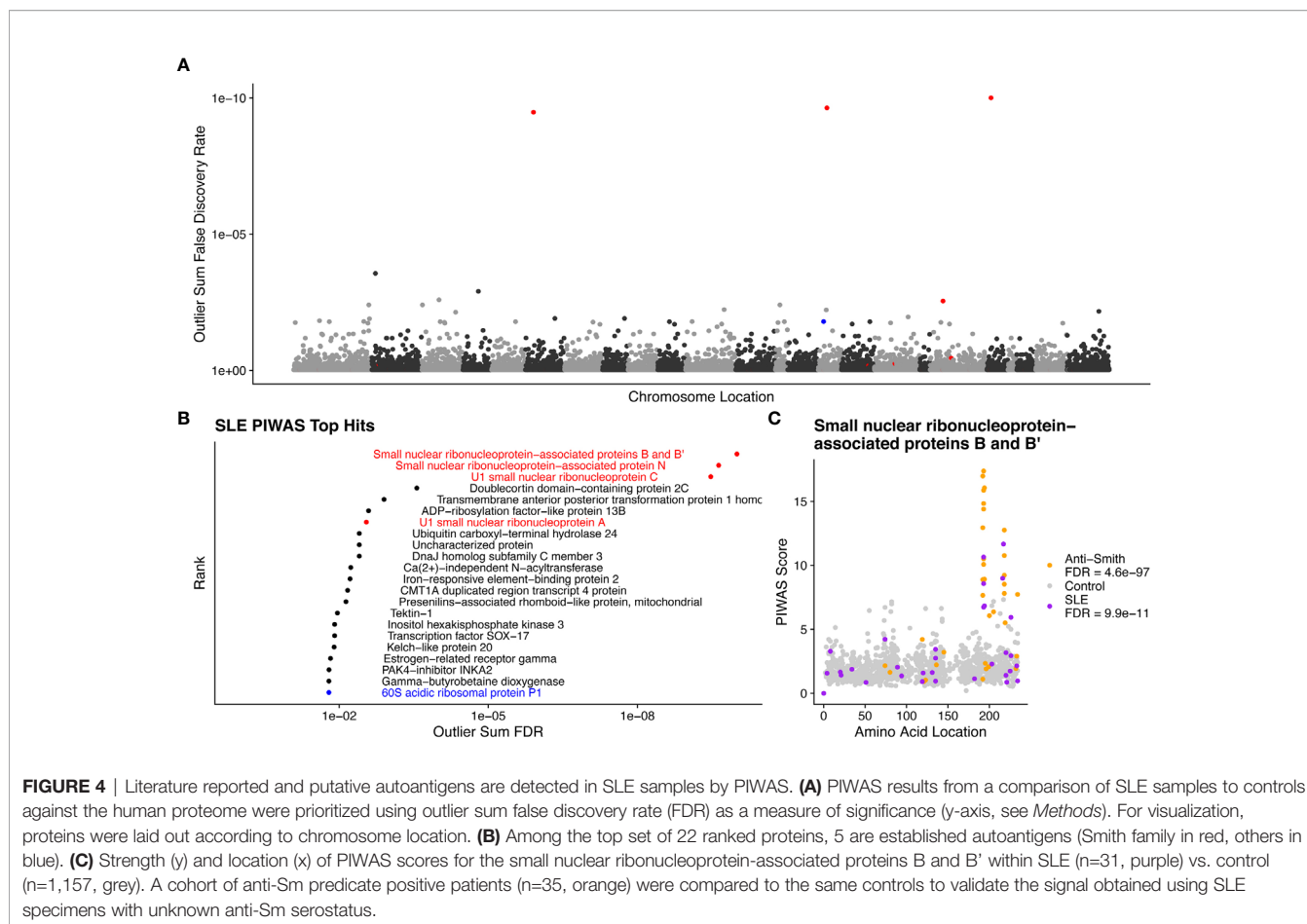
PIWAS Analysis of SERA Datasets From SLE Specimens

We performed PIWAS to identify candidate autoantigens using specimens obtained from SLE patients. PIWAS results from individuals with SLE (n=31) were compared to those from controls (n=1,157) and proteins were ranked based on outlier sum FDR as a measure of significance across the human proteome (21,057 proteins) (**Figures 4A, B**). The highest scoring 22 proteins had outlier sum FDRs ranging from 1.6e-2 to 9.9e-11 and included multiple established autoantigens. Four Smith complex antigens were among the top seven hits with small nuclear ribonucleoprotein-associated proteins B and B' exhibiting the highest significance (outlier sum FDR = 9.9e-11). In addition, 60S acidic ribosomal protein P1, another known SLE autoantigen (20, 38), was highly significant. Multiple highly significant epitopes were evident within nuclear ribonucleoprotein-associated proteins B and B' (**Figure 4C, Table 1**). The most significant enrichments occurred at two different locations near the C-terminus.

PIWAS in an Independent Cohort of Smith Antigen Positive Subjects

To investigate the ability of PIWAS to identify Smith antigens in an independent cohort positive for anti-Sm using validated clinical tests, we applied PIWAS to a cohort of 35 Smith antigen positive samples. In this anti-Sm seropositive cohort, PIWAS again clearly identifies Smith antigens at the top of the ranked list of antigens (**Table 2**). The dominant C-terminal, anti-Sm epitope was identical between the two independent cohorts. The statistical significance within the second cohort is greatly increased relative to the general SLE cohort as might be expected, given the 100% seroprevalence of anti-Smith within this second specimen set. The unbiased identification of known SLE





autoantigens in independent cohorts validates the ability of PIWAS to identify shared autoantigens in a data-driven way.

PIWAS Analysis of SERA Datasets From Sjogren's Syndrome Specimens

We further validated the capability of PIWAS to detect autoantigens in a separate, highly characterized cohort of individuals with Sjogren's syndrome. From a PIWAS analysis comparing individuals with Sjogren's (n=91) to controls (n=1,157), the top ranked antigens across the human proteome (21,057 proteins) included multiple established autoantigens: Lupus La protein (also known as SSB); multiple keratin associated proteins; and Histone H3-like centromeric protein A (CENPA) (**Figures 5A, B**). All of these prominent autoantigens were significant with an outlier sum FDR < 1e-9. To validate the specific reactivities identified within our Sjogren's cohort, we compared PIWAS scores for SSB to predicate SSB status. 16 of 18 individuals with SSB PIWAS signal > 4 were predicate SSB positive (**Figure 5C**, Pearson's chi-squared test p = 0.01). Using an ELISA, we validated that 6 of 8 individuals with high CENPA PIWAS values had reactivity to recombinant CENPA protein (**Figure 5D**, Pearson's chi-squared test p=0.1).

DISCUSSION

We demonstrate the utility of a general and scalable methodology to identify serological antigens within arbitrary proteomes using Protein-based Immunome Wide Association Studies (PIWAS). The power of PIWAS derives from cohort-based statistical analyses within large datasets of antibody-binding epitopes. PIWAS analyzes the enrichments of 5mers and 6mers that have been selected for binding to individual antibody repertoires across a complete proteome. We show that the kmer enrichment space demonstrates specific and enriched signals compared to the unselected libraries. Further, the PIWAS values across the human proteome show greater enrichment in SLE patients compared to control samples. Using synthetic data, we found that PIWAS has power to detect significant antigens at a signal of only 60% of the signal of a known autoantigen. When applied to experimental datasets from SLE cases and controls, PIWAS ranks SLE-specific Smith antigens highly in a proteome-wide search of candidate antigens. We validated the epitopes from this antigen family using a cohort of anti-Sm autoantibody positive patients. Finally, we identified and validated autoantigenic signals to SSB, CENPA, and keratin proteins in a cohort of individuals with Sjogren's syndrome.

TABLE 1 | Dominant epitopes for highest scoring antigens from SLE PIWAS.

Protein Name	Outlier Sum FDR	Dominant epitope(s)
Small nuclear ribonucleoprotein-associated proteins B and B'	9.9E-11	GGPSQQVMTPQ, PGMRPPMGPPM
Small nuclear ribonucleoprotein-associated protein N	2.3E-10	GGPSQQVMTPQ, PGMRPPPPGI
U1 small nuclear ribonucleoprotein C	3.3E-10	GMRPPMGGHMP
Doublecortin domain-containing protein 2C	0.00027	IKPWVHCDINV, YWKSPPVPSEV
Transmembrane anterior posterior transformation protein 1 homolog	0.0013	LLQPAQVCDIL
ADP-ribosylation factor-like protein 13B	0.0026	IASVIIENEGK
U1 small nuclear ribonucleoprotein A	0.0028	PPGMIPPPGLA, PGMIPPPGLAP
Ubiquitin carboxyl-terminal hydrolase 24	0.0039	None
Uncharacterized protein	0.0039	None
DnaJ homolog subfamily C member 3	0.0039	None
Ca(2+)-independent N-acyltransferase	0.0058	LIEGNCEHFVN
Iron-responsive element-binding protein 2	0.006	None
CMT1A duplicated region transcript 4 protein	0.0068	YVITYTSQTVKR, RLIEKSKTREL, SSKSSGKAVFR
Presenilins-associated rhomboid-like protein, mitochondrial	0.0073	GRRFNFFIQK
Tektin-1	0.011	KKLEQRLLEEVD, NSVSLDWDLF
Inositol hexakisphosphate kinase 3	0.012	YDGPDPGYIFG
Transcription factor SOX-17	0.012	QSPPEALPC, MGLPYQGHDSG
Kelch-like protein 20	0.013	None
Estrogen-related receptor gamma	0.015	None
PAK4-inhibitor INKA2	0.016	MDCYLRLKQE, LQDQMNMMGA, TKFSPHRVCG
Gamma-butyrobetaine dioxygenase	0.016	TTGKLSFHTDY, DYCDFSVQSKH
60S acidic ribosomal protein P1	0.016	MGFGLFD

TABLE 2 | Dominant PIWAS epitopes for top antigens from anti-Smith seropositive specimens.

Protein Name	Outlier Sum FDR	Dominant epitope(s)
Small nuclear ribonucleoprotein-associated protein N	1.1e-98	PGMRPPPPGIR
Small nuclear ribonucleoprotein-associated proteins B and B'	4.6e-97	PGMRPPMGPPM
U1 small nuclear ribonucleoprotein A	1.2e-67	PPGMIPPPGLA
U1 small nuclear ribonucleoprotein C	1.3e-47	PGMMPVGPAPG

Previous strategies to proteome-scale antigen identification rely on wet lab approaches that require *a priori* knowledge of the target proteome when the assay is performed (1–6). In contrast, the use of random peptide library data with PIWAS enables analyses against arbitrary proteomes. In addition to the reference human proteome utilized here, the same SERA data can be reanalyzed against proteomes of infectious agents, patient-specific mutations, and splice variants, without performing additional wet lab assays. Indeed, we have identified previously validated epitopes for multiple bacterial, viral, and fungal infectious diseases using this method [data not shown].

PIWAS is an immunological analog to widely employed genome-wide association studies (GWAS) that employ statistical association of gene variants in large disease and control cohorts to identify disease-associated loci. Like GWAS, PIWAS employs a data-driven statistical approach to scan entire genomes and proteomes for statistically significant differences between case and control cohorts. Advancements in GWAS methods such as burden testing has enabled multiple variants within a single gene to be collapsed, thereby increasing the power to detect disease-associated genes (39, 40). Similarly, PIWAS scans each protein to find a maximum signal and allows for the contributions of multiple distinct epitopes to identify candidate antigens associated with disease. By leveraging the outlier sum statistic (37), we are able to further highlight antigens with signals that are strong, but present in only a subset of the

patient population, or derive from unique epitopes within the same antigen.

Just as GWAS must consider a variety of biological and technical limitations, effective PIWAS must consider and address pre-assay, assay, and post-assay factors that can impact performance. The most significant pre-assay issues relate to the selection of cohorts for disease and control populations. Our analyses using synthetic data demonstrated that magnitude and prevalence of autoantigenic signal affects the ability of PIWAS to prioritize antigens. Thus, clean case and control cohorts are more likely to yield genuine autoantigens. In this study we were able to detect known antigens using a small cohort of SLE cases. As the cohort size grows, we anticipate even greater power to identify known and novel autoantigens.

Application of PIWAS to a cohort of SLE subjects identified known autoantigens, with 5 of 16 of the highest ranking hits across the entire human proteome being validated and clinically significant autoantigens. In particular, Smith antigens stood out as top hits in the SLE analysis. To validate this finding, we analyzed specimens from a second independent cohort of patients that tested positive for anti-Sm using clinical predicate tests. We found that the anti-Sm positive cohort exhibited stronger reactivity against the same Sm antigens and epitopes as expected compared to the less homogeneous SLE discovery cohort. PIWAS identified an anti-Sm epitope occurring within a proline rich region in agreement with multiple prior studies (20, 41).

To test the generalizability of the PIWAS method, we identified multiple validated antigens in a cohort of 91 subjects with Sjogren's disease, including Lupus La, CENPA, and keratins. Notably, the immunodominant region seen in Sjogren's subjects in La/SSB, located in the region 349-408, has been previously described by Tziofas et al. (amino acids 349-368), validating the PIWAS epitope finding (42). While literature on the additional putative ranked antigens identified is limited for Sjogren's disease, literature regarding the role of these putative antigens (although not necessarily autoantibodies against them) in other diseases is suggestive of their potential involvement in autoimmune or inflammatory processes. For example, prostaglandin receptor E2 EP2 subtype has been associated with both pro-inflammatory and regulatory roles in SLE and rheumatoid arthritis (43-45), suggesting it could have similar effects in Sjogren's disease. Presenilins-associated rhomboid-like protein, mitochondrial (PARL) is a mitochondrial protease that cleaves PINK1, implicated in Parkinson's disease and decreased activity could be associated with increased mitophagy through Parkin (46). Interestingly, this antigen was also found independently to be significant in the SLE cohort. Finally, double C2 domain-containing protein β (Doc2b) protects β -cells against inflammatory damage in mice models of diabetes through reducing apoptotic stress (47).

SERA's large random library, the use of a large control set in PIWAS, flexibility to arbitrary proteins, and the incorporation of multiple epitopes per antigen together power the discovery capabilities of the platform relative to peptide arrays or directed phage display libraries. While it is difficult to make direct comparisons to other platforms due to differences in cohorts (demographics, disease states, clinical characterization), technologies, and methods, we can compare reported results in observed epitopes and antigens. In our SLE cohorts, we have identified many of the autoantigens detected using other peptide platforms, including Smith antigens and anti-ribosomal P (48, 49). Although signal for other characterized autoantigens (including histone H2B and topoisomerase) did not meet statistical criteria for positivity in the SLE study, outlier signals were observed for these antigens in some individual subjects (data not shown). We identify many autoantigens with characterized relevance in Sjogren's, including Lupus La protein, CENPA, and keratin associated proteins. We report additional antigens that were not previously characterized in these disease states, including TRIM33 and PARL. We validate the legitimacy of many proposed autoantigens based on prior literature, but we also note that some important autoantigens were not detected [i.e. Ro and ANA (50, 51)] due to a variety of factors. For example, Lupus La was identified as a significant antigen in the Sjogren's cohort but not in the SLE cohort, even though it is well characterized in both diseases. In this case, possible explanations likely include the size and composition of the disease cohorts. For other antigens that are detected only by protein arrays, non-linear (structural) epitopes are likely to have a significant limitation on detection by PIWAS.

In all cases, the presence of autoantibodies to these proteins and their functional significance in SLE or Sjogren's would need to be

validated, which is beyond the scope of this study, but they demonstrate the potential use of PIWAS for novel antigen discovery. PIWAS ranks antigens based on the maximum signal observed across a cohort, however it is not always possible to determine which antigens are biologically significant due to sequence similarity between proteins. Therefore, antigens ranked highly in PIWAS should be considered candidate antigens, and orthogonal experimental validation is generally necessary to establish a *bona fide* antigen. If these candidate autoantigens are validated, they could be incorporated into multi-analyte autoantigen panels for diagnostic or prognostic purposes. SLE has a notable sex bias (9 female:1 male), which is roughly reflected in our cohort. When we adjust the control population to match this sex distribution, the top antigens remain similar, though significance decays due to the decreased control population size. Generally, larger control populations for PIWAS analyses yield superior performance.

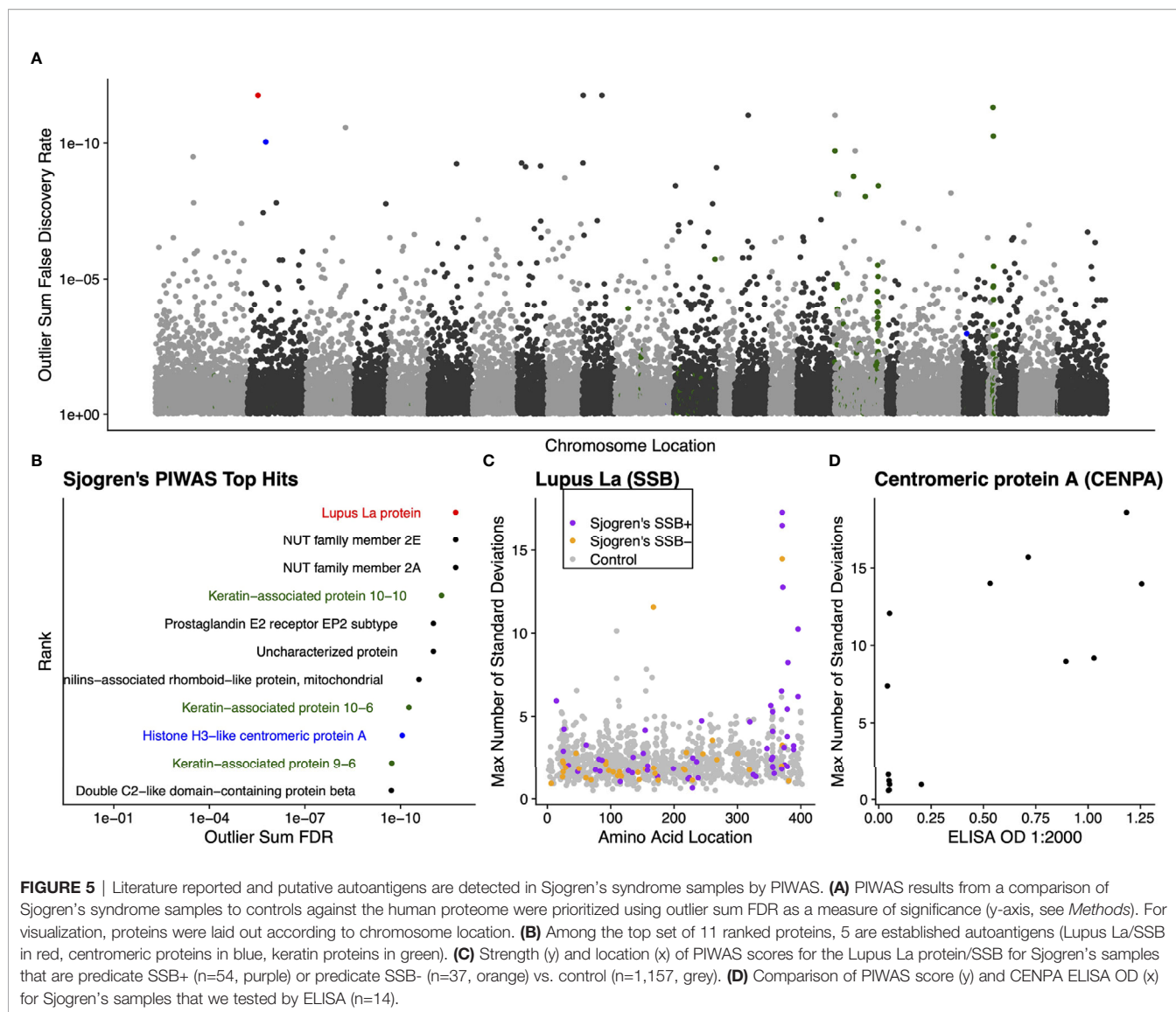
Although many antibody epitopes contain a linear or contiguous segment of at least 5 amino acids (33), those with purely conformational epitopes or mimotopes may not be identified using PIWAS. PIWAS, as presented, is limited to identifying linear epitopes at a proteome scale. We have found that motifs (which allow degenerate positions) can more accurately capture mimotope specificities while maintaining some linear mapping (52). Thus, we are developing PIWAS that leverages motif patterns identified by IMUNE (26). As presented, our method leverages 12mer data generated by the SERA platform, but is extensible to arbitrary input sequence lengths, kmer lengths, and window sizes. PIWAS draws statistical power from the large control cohort (which is publicly available): both to accurately estimate the mean and standard deviation for every kmer and to maintain high specificity while performing proteome-scale discovery analysis. Furthermore, the current method uses the maximum signal observed within the protein sequence for a particular patient, but some antigens have multiple antibody epitopes (53). The use of multiple signals within a protein is another avenue of development to improve both sensitivity and specificity of PIWAS.

In conclusion, we developed PIWAS to enable robust, proteome-wide, cohort-based antigen discovery. PIWAS analyzes the datasets resulting from random peptide library selections against case and control cohorts (e.g., SERA) to discover shared candidate antigens, regardless of whether the epitopes therein are public or private. Since SERA employs random libraries, PIWAS can be applied to multiple proteomes utilizing the same physical assay. As the size of case and control datasets continue to increase, PIWAS may uncover previously undiscovered antigens with potential utility in diagnostic and therapeutic applications. Finally, PIWAS may be useful to investigate, in an unbiased manner, the association of autoantigens, human pathogens, and commensal organisms with human disease.

SUPPLEMENTAL METHODS

Serum Epitope Repertoire Analysis (SERA)

Development and preparation of the *Escherichia coli* random 12-mer peptide display library (diversity 8×10^9) has been described



previously (26). SERA was performed as described (26). Briefly, serum was diluted 1:25 and incubated for 1 hr with a 10-fold oversampling of the library (8×10^{10} cells/well) in a 96-well plate format at 4°C with orbital shaking (800 rpm) during which time serum antibodies bind to peptides on the bacterial surface that mimic their cognate antigens. Cells were then collected by centrifugation (3500 rcf x 7 min), the supernatant was removed, and the cell pellets were washed by resuspending in 750 μ L PBS + 0.05% Tween-20 (PBST). The cells were again collected by centrifugation (3500 rcf x 7 min) and the supernatant was removed. Cell pellets were resuspended in 750 μ L PBS and mixed thoroughly with 50 μ L Protein A/G Sera-Mag SpeedBeads (GE Life Sciences, 17152104010350) (6.25% the beads' stock concentration). The plate was incubated for one hour at 4°C with orbital shaking (800 rpm). Bead-bound cells were captured in the plate using a Magnum FLX 96-ring magnet (Alpaqua, A000400) until all beads were separated. Unbound cells in the supernatant were removed by gentle pipetting, leaving

only those cells bound to A/G beads. Beads were washed 5X by removing from the magnet, resuspending in 750 μ L PBST, and then returning to the magnet. The supernatant was removed by gentle pipetting after the beads were securely captured. Cells were resuspended in 750 μ L LB with 34 μ g/mL chloramphenicol and 0.2% wt/vol glucose directly in the 96-deep-well plate and grown overnight with shaking (300 rpm) at 37°C.

Amplicon Library Preparation for Sequencing

After growth, cells were collected by centrifugation (3500 rcf for 10 min) and the supernatant was discarded. Plasmids encoding the selected peptides were isolated in 96-well format using the Montage Plasmid Miniprep_{HTS} Kit (MilliPore, LSKP09604) on a Multiscreen_{HTS} Vacuum Manifold (MilliPore, MSVMHTS00) following the "Plasmid DNA—Full Lysate" protocol in the product literature. For amplicon preparation, two rounds of PCR were employed; the first round amplifies the variable "X12" peptide region of the plasmid DNA. The second round

barcodes each patient amplicon library with sample-specific indexing primers for data demultiplexing after sequencing. KAPA HiFi HotStart ReadyMix (KAPA Biosystems, KK2612) was used as the polymerase master mix for all PCR steps. Plasmids (2.5 μ L/well) were used as template for a first round PCR with 12.5 μ L of KAPA ReadyMix and 5 μ L each of 1 μ M forward and reverse primers. The primers (Integrated DNA Technologies) contain annealing regions that flank the X12 sequence (indicated in bold) and adapter regions specific to the Illumina index primers used in the second round PCR.

Forward primer: TCGTCGGCAGCGTCAGATGTG TATAAGAGACAGVBHDV**CCAGTCTGGCCAGGG**

Reverse primer: GTCTCGTGGGCTCGGAGATG TGTATAAGAGACAGGT**GATGCCGTAGTACTGG**

A series of five degenerate bases in the forward primer, VBHDV (following IUPAC codes), provide base diversity for the first five reads of the sequencing on the NextSeq platform. The five base pairs were designed to be non-complementary to the template to avoid bias during primer annealing. To reduce non-specific products, a touchdown PCR protocol was used with an initial annealing temperature of 72°C with a decrease of 0.5°C per cycle for 14 cycles, followed by 10 cycles with annealing at 65°C. The 25 μ L primary PCR product was purified using 30 μ L Mag-Bind TotalPure NGS Beads (Omega Bio-Tek, M1378-02) according to the manufacturer's protocol. The second round PCR (8 cycles, 70°C annealing temperature) was performed using Nextera XT index primers (Illumina, FC-131-2001) which introduce 8 base pair indices on the 5' and 3' termini of the amplicon for data demultiplexing of each sample screened. The PCR 1 product (5 μ L) was used as a template for the second PCR with 5 μ L each of forward and reverse indexing primers, 5 μ L PCR grade water and 25 μ L of KAPA ReadyMix. The PCR product (50 μ L) was cleaned up with 56 μ L Omega Mag-Bind TotalPure NGS Beads per reaction. A 96-well quantitation was performed using the Qubit dsDNA High Sensitivity assay (Invitrogen, Q32851) adapted for a microplate fluorimeter (Tecan SPECTRAFlour Plus) measuring fluorescence excitation at 485 nm and emission at 535 nm. Positive (100 ng) and negative (0 ng) controls, included with the Qubit kit, were added to the plate as standards along with 2 μ L of each PCR product diluted 1:100 for quantitation. The fluorescence data were used to calculate DNA concentration in each well based on the kit standards. To normalize the DNA and achieve equal loading of each patient sample on NGS, the DNA in each well was diluted with Tris HCl (pH 8.5, 10 mM) to 4 nM and an equal volume from each well was pooled in a Lo-Bind DNA tube for sequencing.

The sample pool was prepared for sequencing according to specifications of the Illumina NextSeq 500. Due to the low diversity in the adapter regions of the amplicon after the first five bases, PhiX Run Control (Illumina, FC-110-3001) was included at 40% of the final DNA pool. The pool was sequenced using a High Output v2, 75 cycle kit (Illumina, FC-404-2005).

Naïve Library Sequencing

An aliquot of the naïve X12 library representing 10-fold oversampling of the diversity was divided into 10 tubes, and the plasmids were purified and amplicons prepared as described

above. Each prep was barcoded with a unique set of indices and sequenced on the NextSeq 500 to yield approximately 400 million unique sequences.

CENPA ELISA

Briefly, recombinant proteins CENPA (Origene, Cat# TP301602), or NY-ESO-1 (used as a control) (Origene, Cat# TP313318) were coated onto flat bottom, 96 well plates (Nunc MaxiSorp), at 0.5 μ g/ml in PBS, 50 μ L per well at 4°C overnight. After washing with PBS containing 0.1% Tween 20, plates were blocked with PBS containing 5% non-fat milk for 2 hours at room temperature. Plates were then incubated with 100 μ L serum diluted 1/2000 in blocking buffer for 1 hour at room temperature. Following washing, plates were incubated with 100 μ L HRP-goat anti-human IgG (Jackson ImmunoResearch) secondary antibody diluted 1/10,000 in blocking buffer for 1 hour at room temperature. Plates were washed, and the reaction was developed with 3,3',5,5'-teramethylbenzidine substrate solution (ThermoFisher) for 1-10 minutes and stopped with an equal volume of 1M HCL. The absorbance at 450 nm was measured on a Tecan Spectrafluor plus plate reader.

Cohorts

Control Cohort

Specimens from 1,157 apparently healthy individuals were used as a control cohort.

SLE Cohort

De-identified specimens from 31 individuals diagnosed with SLE, and primarily female (27), were acquired from Proteogenex (9) and BioIVT (22). The mean age within this cohort was 43 years, with a range of 22-72.

Anti-Smith Cohort

Samples from 34 subjects that tested positive for Anti-SM RNP (4) or Anti-Smith (30) antibodies by predicate ANA multiplex testing were obtained from Discovery Life Sciences. Subjects ranged in age from 18 -74, with the majority (26) being female.

Sjogren's Syndrome Cohort

Samples from 91 individuals diagnosed with Sjogren's syndrome were acquired from the Sjogren's International Collaborative Clinical Alliance [SICCA]. Subject ranged in age from 21 to 78 years, with a mean of 52 years. The majority of subjects (81) are female.

Proteome Description

The reference *Homo sapiens* proteome was downloaded from Uniprot (54) on February 28, 2019.

Kmer Enrichment Analysis

We compared the count of unique kmer species vs. enrichment scores for 5 and 6 mers in assays with a random library vs. those incubated with serum. We also compared the distribution of PIWAS values and average PIWAS values across control and SLE samples.

Autoantigen Simulation Experiments

To simulate the effects of changing the magnitude and prevalence of autoantigenic signal, the real PIWAS signal against one of the Smith

antigens in the SLE cohort was selected for use in a series of simulations (P14678: Small nuclear ribonucleoprotein-associated proteins B and B'). For every sample, the PIWAS values were calculated. To simulate different magnitudes of effect, the SLE PIWAS values were multiplied by scaling factors ranging from [0.1,2] and the outlier sum statistics were calculated relative to unscaled control values. To simulate different prevalences of effect, the SLE PIWAS values were divided into "high" (PIWAS > 6) and "low" (PIWAS < 6) values, 1000 random samplings with replacement of the SLE cohort were taken to simulate prevalences of "high" ranging from [0.01, 1], and the outlier sum statistics were calculated relative to unaffected control values.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for

participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

Conceptualization: WH, KK, PD, and JS. Sample screening and analysis: KK and RW. Data curation: WH. Formal analysis: WH. Funding acquisition: PD and JS. Software: WH. Visualization: WH. Writing- Review & Editing: WH, KK, PD, and JS. Writing-Original draft preparation: WH and JS. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We thank the Serimmune team for supporting the development of PIWAS and processing of samples: Minlu Zhang, Jack Reifert, Joel Bozekowski, Burak Himmetoglu, Brian Martinez, Gregory Jordan, Timothy Johnston, Cameron Gable, Steve Kujawa, Elisabeth Baum-Jones. Special thanks to Elizabeth Stewart for editing this manuscript. Sjogren's specimens used in this manuscript are from the Sjogren's International Collaborative Clinical Alliance [SICCA], funded under contract N01 DE-32636 by the National Institute of Dental and Craniofacial Research, with funding support from the National Eye Institute and Office for Research in Women's Health.

REFERENCES

- Zwick C, Pfreundschuh M.SEREX. In: *Encyclopedia of Cancer*. Berlin, Heidelberg: Springer Berlin Heidelberg. (2011) p. 3380–1. doi: 10.1007/978-3-642-16483-5_5252
- Shi YY, Wang HC, Yin YH, Sun WS, Li Y, Zhang CQ, et al. Identification and Analysis of Tumour-Associated Antigens in Hepatocellular Carcinoma. *Br J Cancer* (2005) 92:929–34. doi: 10.1038/sj.bjc.6602460
- Zhou FL, Zhang WG, Chen G, Zhao WH, Cao XM, Chen YX, et al. Serological Identification and Bioinformatics Analysis of Immunogenic Antigens in Multiple Myeloma. *Cancer Immunol Immunother* (2006) 55:910–7. doi: 10.1007/s00262-005-0074-x
- Larman HB, Zhao Z, Laserson U, Li MZ, Ciccio A, Gakidis MAM, et al. Autoantigen Discovery With a Synthetic Human Peptidome. *Nat Biotechnol* (2011) 29:535–41. doi: 10.1038/nbt.1856
- Xu GJ, Kula T, Xu Q, Li MZ, Vernon SD, Ndong'u T, et al. Comprehensive Serological Profiling of Human Populations Using a Synthetic Human Virome. *Science* (80-) (2015) 348:aaa0698. doi: 10.1126/science.aaa0698
- Zandian A, Forsström HA, Häggmark-Månberg A, Schwenk JM, Uhlé M, Nilsson P, et al. Whole-Proteome Peptide Microarrays for Profiling Autoantibody Repertoires Within Multiple Sclerosis and Narcolepsy. *J Proteome Res* (2017) 16:1300–14. doi: 10.1021/acs.jproteome.6b00916
- Richer J, Johnston SA, Stafford P. Epitope Identification From Fixed-Complexity Random-Sequence Peptide Microarrays. *Mol Cell Proteomics* (2015) 14:136–47. doi: 10.1074/mcp.M114.043513
- Hansen LB, Buus S, Schafer-Nielsen C. Identification and Mapping of Linear Antibody Epitopes in Human Serum Albumin Using High-Density Peptide Arrays. *PLoS One* (2013) 8:e68902. doi: 10.1371/journal.pone.0068902
- Buus S, Rockberg J, Forsström B, Nilsson P, Uhlen M, Schafer-Nielsen C. High-Resolution Mapping of Linear Antibody Epitopes Using Ultrahigh-Density Peptide Microarrays. *Mol Cell Proteom* (2012) 11:1790–800. doi: 10.1074/mcp.M112.020800
- Finn OJ. Human Tumor Antigens Yesterday, Today, and Tomorrow. *Cancer Immunol Res* (2017) 5:347–54. doi: 10.1158/2326-6066.CIR-17-0112
- Zhao Z, Ren J, Dai C, Kannapell CC, Wang H, Gaskin F, et al. Nature of T Cell Epitopes in Lupus Antigens and HLA-DR Determines Autoantibody Initiation and Diversification. *Ann Rheum Dis* (2018) 78:380–90. doi: 10.1136/annrheumdis-2018-214125
- 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:e67181. doi: 10.1371/journal.pone.0067181
- Chan LS, Vanderlugt CJ, Hashimoto T, Nishikawa T, Zone JJ, Black MM, et al. Epitope Spreading: Lessons From Autoimmune Skin Diseases. *J Invest Dermatol* (1998) 110:103–9. doi: 10.1046/j.1523-1747.1998.00107.x
- Didona D, Di Zenzo G. Humoral Epitope Spreading in Autoimmune Bullous Diseases. *Front Immunol* (2018) 9:779. doi: 10.3389/fimmu.2018.00779
- Rosen A, Casciola-Rosen L. Autoantigens in Systemic Autoimmunity: Critical Partner in Pathogenesis. *J Internal Med* (2009) 265:625–31. doi: 10.1111/j.1365-2796.2009.02102.x
- Zaenker P, Gray ES, Ziman MR. Autoantibody Production in Cancer-The Humoral Immune Response Toward Autologous Antigens in Cancer Patients. *Autoimmun Rev* (2016) 15:477–83. doi: 10.1016/j.autrev.2016.01.017
- Yaniv G, Twig G, Shor DBA, Furer A, Sherer Y, Mozes O, et al. A Volcanic Explosion of Autoantibodies in Systemic Lupus Erythematosus: A Diversity of 180 Different Antibodies Found in SLE Patients. *Autoimmun Rev* (2015) 14:75–9. doi: 10.1016/j.autrev.2014.10.003
- Riemekasten G, Hahn BH. Key Autoantigens in SLE. *Rheumatol (Oxford)* (2005) 44:975–82. doi: 10.1093/rheumatology/keh688
- Ching KH, Burbelo PD, Tipton C, Wei C, Petri M, Sanz I, et al. Two Major Autoantibody Clusters in Systemic Lupus Erythematosus. *PLoS One* (2012) 7. doi: 10.1371/journal.pone.0032001
- Dema B, Charles N. Autoantibodies in SLE: Specificities, Isotypes and Receptors. *Antibodies* (2016) 5:2. doi: 10.3390/antib5010002

21. Kalinina O, Louzoun Y, Wang Y, Utset T, Weigert M. Origins and Specificity of Auto-Antibodies in Sm+ SLE Patients. *J Autoimmun* (2018) 90:94–104. doi: 10.1016/j.jaut.2018.02.008
22. Deshmukh US, Bagavant H, Lewis J, Gaskin F, Shu MF. Epitope Spreading Within Lupus-Associated Ribonucleoprotein Antigens. *Clin Immunol* (2005) 117:112–20. doi: 10.1016/j.clim.2005.07.002
23. James JA, Harley JB. Linear Epitope Mapping of an Sm B/B' Polypeptide. *J Immunol* (1992) 148:2074–9.
24. Gottenberg JE, Mignot S, Nicaise-Rolland P, Cohen-Solal J, Aucouturier F, Goetz J, et al. Prevalence of Anti-Cyclic Citrullinated Peptide and Anti-Keratin Antibodies in Patients With Primary Sjögren's Syndrome. *Ann Rheum Dis* (2005) 64:114–7. doi: 10.1136/ard.2003.019794
25. Shen L, Suresh L. Autoantibodies, Detection Methods and Panels for Diagnosis of Sjögren's Syndrome. *Clin Immunol* (2017) 182:24–9. doi: 10.1016/j.clim.2017.03.017
26. Pantazes RJ, Reifert J, Bozekowski J, Ibsen KN, Murray JA, Daugherty PS. Identification of Disease-Specific Motifs in the Antibody Specificity Repertoire Via Next-Generation Sequencing. *Sci Rep* (2016) 6:30312. doi: 10.1038/srep30312
27. Bozekowski JD, Graham AJ, Daugherty PS. High-Titer Antibody Depletion Enhances Discovery of Diverse Serum Antibody Specificities. *J Immunol Methods* (2018) 455:1–9. doi: 10.1016/j.jim.2018.01.003
28. Getz JA, Schoep TD, Daugherty PS. Peptide Discovery Using Bacterial Display and Flow Cytometry. *Methods Enzymol* (2012) 75–97. doi: 10.1016/B978-0-12-396962-0.00004-5
29. Schreiber A, Humbert M, Benz A, Dietrich U. 3d-Epitope-Explorer (3dex): Localization of Conformational Epitopes Within Three-Dimensional Structures of Proteins. *J Comput Chem* (2005) 26:879–87. doi: 10.1002/jcc.20229
30. Moreau V, Granier C, Villard S, Laune D, Molina F. Discontinuous Epitope Prediction Based on Mimotope Analysis. *Bioinformatics* (2006) 22:1088–95. doi: 10.1093/bioinformatics/btl012
31. Mayrose I, Penn O, Erez E, Rubinstein ND, Shlomi T, Freund NT, et al. Peptote: Epitope Mapping From Affinity-Selected Peptides. *Bioinformatics* (2007) 23:3244–6. doi: 10.1093/bioinformatics/btm493
32. Huang J, Gutteridge A, Honda W, Kanehisa M. Mimox: A Web Tool for Phage Display Based Epitope Mapping. *BMC Bioinf* (2006) 7:451. doi: 10.1186/1471-2105-7-451
33. Kringelum JV, Nielsen M, Padkjær SB, Lund O. Structural Analysis of B-cell Epitopes in Antibody: Protein Complexes. *Mol Immunol* (2013) 53:24–34. doi: 10.1016/j.molimm.2012.06.001
34. Sun J, Xu T, Wang S, Li G, Wu D, Cao Z. Does Difference Exist Between Epitope and non-Epitope Residues? Analysis of the Physicochemical and Structural Properties on Conformational Epitopes From B-cell Protein Antigens. *Immunome Res* (2011) 7:3:1.
35. Sun P, Ju H, Liu Z, Ning Q, Zhang J, Zhao X, et al. Bioinformatics Resources and Tools for Conformational B-cell Epitope Prediction. *Comput Math Methods Med* (2013) 2013:1–11. doi: 10.1155/2013/943636
36. 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:e0217668. doi: 10.1371/journal.pone.0217668
37. Tibshirani R, Hastie T. Outlier Sums for Differential Gene Expression Analysis. *Biostatistics* (2007) 8:2–8. doi: 10.1093/biostatistics/kxl005
38. Gerli R, Caponi L. Anti-Ribosomal P Protein Antibodies. *Autoimmunity* (2005) 38:85–92. doi: 10.1080/08916930400022699
39. Madsen BE, Browning SR. A Groupwise Association Test for Rare Mutations Using a Weighted Sum Statistic. *PLoS Genet* (2009) 5:e1000384. doi: 10.1371/journal.pgen.1000384
40. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-Variant Association Analysis: Study Designs and Statistical Tests. *Am J Hum Genet* (2014) 95:5–23. doi: 10.1016/j.ajhg.2014.06.009
41. Sundar K, Jacques S, Gottlieb P, Villars R, Benito ME, Taylor DK, et al. Expression of the Epstein-Barr Virus Nuclear Antigen-1 (EBNA-1) in the Mouse can Elicit the Production of anti-dsDNA and anti-Sm Antibodies. *J Autoimmun* (2004) 23:127–40. doi: 10.1016/j.jaut.2004.06.001
42. Tzioufas AG, Yiannaki E, Sakarellos-Daitsiotis M, Routsias JG, Sakarellos C, Moutsopoulos HM. Fine Specificity of Autoantibodies to La/SSB: Epitope Mapping, and Characterization. *Clin Exp Immunol* (1997) 108:191–8. doi: 10.1046/j.1365-2249.1997.d01-1003.x
43. Fabricius D, Neubauer M, Mandel B, Schütz C, Viardot A, Vollmer A, et al. Prostaglandin E 2 Inhibits Ifn- α Secretion and Th1 Costimulation by Human Plasmacytoid Dendritic Cells Via E-Prostanoid 2 and E-Prostanoid 4 Receptor Engagement. *J Immunol* (2010) 184:677–84. doi: 10.4049/jimmunol.0902028
44. Akaogi J, Nozaki T, Satoh M, Yamada H. Role of PGE2 and EP Receptors in the Pathogenesis of Rheumatoid Arthritis and as a Novel Therapeutic Strategy. *Endocr Metab Immune Disord - Drug Targets* (2012) 6:383–94. doi: 10.2174/187153006779025711
45. McCoy JM, Wicks JR, Audoly LP. The Role of Prostaglandin E2 Receptors in the Pathogenesis of Rheumatoid Arthritis. *J Clin Invest* (2002) 110:651–8. doi: 10.1172/jci15528
46. Greene AW, Grenier K, Aguilera MA, Muise S, Farazifard R, Haque ME, et al. Mitochondrial Processing Peptidase Regulates PINK1 Processing, Import and Parkin Recruitment. *EMBO Rep* (2012) 13:378–85. doi: 10.1038/embor.2012.14
47. Aslmy A, Oh E, Olson EM, Zhang J, Ahn M, Moin ASM, et al. Doc2b Protects B-Cells Against Inflammatory Damage and Enhances Function. *Diabetes* (2018) 67:1332–44. doi: 10.2337/db17-1352
48. Zhang Q, Li X, Li H, Hu Q, Lin J, Li Y, et al. Detection of Epitopes in Systemic Lupus Erythematosus Using Peptide Microarray. *Mol Med Rep* (2018) 17:6533–41. doi: 10.3892/mmr.2018.8710
49. Wang L, Hao C, Deng Y, Liu Y, Hu S, Peng Y, et al. Screening Epitopes on Systemic Lupus Erythematosus Autoantigens With a Peptide Array. *Oncotarget* (2017) 8:85559–67. doi: 10.18632/oncotarget.20994
50. Li QZ, Zhou J, Wandstrat AE, Carr-Johnson F, Branch V, Karp DR, et al. Protein Array Autoantibody Profiles for Insights Into Systemic Lupus Erythematosus and Incomplete Lupus Syndromes. *Clin Exp Immunol* (2007) 147:60–70. doi: 10.1111/j.1365-2249.2006.03251.x
51. Utz PJ. Multiplexed Assays for Identification of Biomarkers and Surrogate Markers in Systemic Lupus Erythematosus. *Lupus* (2004) 13:304–11. doi: 10.1191/0961203303lu1017oa
52. Chen WS, Haynes WA, Waitz R, Kamath K, Vega-Crespo A, Shrestha R, et al. Autoantibody Landscape in Patients With Advanced Prostate Cancer. *Clin Cancer Res* (2020) 26:6204–14. doi: 10.1158/1078-0432.ccr-20-1966
53. Zhang L. Multi-Epitope Vaccines: A Promising Strategy Against Tumors and Viral Infections. *Cell Mol Immunol* (2018) 15:182–4. doi: 10.1038/cmi.2017.92
54. UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Res* (2017) 45: D158–69. doi: 10.1093/nar/gkw1099

Conflict of Interest: The authors declare the following competing interests: ownership of stocks or shares at Serimmune, paid employment at Serimmune, board membership at Serimmune, and patent applications on behalf of Serimmune.

Copyright © 2021 Haynes, Kamath, Waitz, Daugherty and Shon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.