AWARD NUMBER: W81XWH-19-1-0639

TITLE: Antigenome Signatures as Biomarkers for Subtyping Disease Heterogeneity in Lupus Patients

PRINCIPAL INVESTIGATOR: Thomas F. Tedder

CONTRACTING ORGANIZATION: Duke University, Durham, NC

REPORT DATE: JANUARY 2023

TYPE OF REPORT: FINAL

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JANUARY 2023	FINAL	30Sep2019 - 29SEPT2022

4. TITLE AND SUBTITLE

Antigenome Signatures as Biomarkers for Subtyping Disease Heterogeneity in Lupus Patients

5a. CONTRACT NUMBER
W81XWH-19-1-0639

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Thomas F. Tedder

E-Mail: thomas.tedder@duke.edu

5d. PROJECT NUMBER
0011351690

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Duke University
2200 W. Main Street
Suite 710
Durham, NC 27708-4677

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Background: Systemic lupus erythematosus (SLE) is a heterogeneous disease for which diagnosis relies heavily on the presence of serum anti-nuclear autoantibodies (ANA). Although 70 years of research has crowned immunofluorescence staining of HEp-2 cells the "golden standard" for ANA detection, the unique and overlapping autoantibody specificities that can jointly produce 30 unique staining patterns remain generally unknown. Moreover, standard clinical assay kits in a recent study failed to detect measurable ANA in 5-24% of patients with established SLE. While technical shortcomings of ANA assay kits may underlie the absence of measurable ANA titers in some cases, there remains great disagreement among experts in the field regarding measures needed for assay optimization and standardization. Proper ANA detection is critical for diagnosis and treatment as ANA seropositive patients demonstrate increased responsiveness to therapy and are therefore preferentially recruited into clinical trials. Thus, a better molecular understanding of ANA reactivity and the patient subgroups that these autoantibodies define is acutely needed.

Focus Areas: New biomarkers and technologies for identifying and subtyping SLE patients are needed to better classify and track disease heterogeneity, identify progressive stages of disease over time, and provide personalized medicine strategies for patient treatment and a better understanding of lupus disease mechanisms.

Hypothesis/Objective: ANA expression is fundamental to current paradigms for diagnosing and classifying SLE patients and may contribute to disease pathogenesis. We therefore hypothesize that ANA assay results reflect complex mixtures of hundreds of autoantibody specificities that may individually or selectively serve as diagnostic, prognostic, or treatment biomarkers for different patient cohorts or phases of disease.

Specific Aims: Using a unique state-of-the-art technology that simultaneously discovers, identifies, and maps autoantibody targets across the human genome, we will test this hypothesis using well-characterized/well-annotated serum samples from SLE patients and evaluate the disease specificity of autoantibodies in two aims:

Aim 1. Comprehensively identify the diversity of molecular targets recognized by ANA standard sera and SLE autoantibodies. Highly characterized ANA standard sera as well as ANA+ and ANA- SLE patient's sera will be screened against a novel genome-wide protein display library. Representing ~19,600 human proteins, the library displays ~5 million domain-sized protein fragments in solution, thereby preserving native conformations. With its advanced deep-sequencing platform and a novel bioinformatics pipeline, this technology enables the simultaneous quantitative and reproducible mapping of the ~60 trillion antibody molecules present in 1.0 microliter of human serum.

Aim 2. Validate the disease and cohort specificity of SLE autoantigen signatures. The ~700 million autoantigen data points generated in Aim 1 will be extensively analyzed to identify autoantigen signatures for each serum sample. Individual- and cohort-specific autoantigen specificities will be correlated with each sera's ANA reactivity, ANA staining pattern, and patient disease features. Disease- and subset-relevant autoantigens will be further evaluated in the context of an extensive autoantigen signature database that currently represents 136 autoimmune disease and healthy control sera.

15. SUBJECT TERMS
None listed.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRDC
U	U	U	UU	10	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

	Page
1. Introduction	3
2. Keywords	3
3. Accomplishments	4
4. Impact	7
5. Changes/Problems	7
6. Products	8
7. Participants & Other Collaborating Organizations	8
8. Special Reporting Requirements	10
9. Appendices	10

Introduction

Anti-nuclear antibodies (ANAs) are autoantibodies directed to macromolecular components of the cell nucleus and remain a fundamental basis for defining, classifying and diagnosing systemic lupus erythematosus (SLE), a prototypic systemic autoimmune disease. Pathophysiologic differences between ANA seropositive and seronegative patients is a major issue in assessing classification and clinical disease activity, with data strongly suggesting that serological findings affect disease severity treatment outcomes. Because of the relationship between serological status and treatment outcome, current clinical trials for new agents to treat SLE predominantly focus on enrolling only ANA and/or anti-DNA or anti-Sm antibody positive patients. Thus, ANA positivity is increasingly required for patients to enter clinical trials and to receive new therapies that are FDA approved for seropositive patients only. Understanding the complexity of immune targets in SLE also remains a significant challenge. ANA test results are dependent on complex autoantibody mixtures in each patient as reflected in the 30 distinct ANA staining patterns (www.ANApatterns.org) considered in evaluations of patients with SLE and other autoimmune disorders. Obtaining a more complete molecular understanding of each SLE patient's autoantibody specificities and their relationship to disease manifestations and activity over time remains an urgent challenge. To address the need for more informative biomarkers, these studies have tested the hypothesis that ANA assay results reflect complex mixtures of autoantibody specificities that may serve as diagnostic, prognostic, and theranostic biomarkers for different phases of SLE disease. Understanding serum ANA complexity as a fundamental basis for both defining and diagnosing SLE has represented our focus throughout the project; there have not been any significant changes in the project nor its direction.

2. Keywords

ANA - anti-nuclear antibodies
Antigenome – the repertoire of autoantigens encoded by the human genome
SLE – systemic lupus erythematosus
Autoantibodies
Biomarkers
Antigenome Platform – the assay we have constructed to assess autoantibody targets

3. Accomplishments

Major goals of the project:

There are two major goals of the project.

Aim 1. Identify the diversity of molecular targets recognized by ANA standard sera and SLE autoantibodies. Using the genome-wide human protein expression library and the Antigenome Platform, a total of 70 serum samples will be screened including ANA standard sera and highly annotated sera from documented ANA-positive and ANA-negative SLE patients.

Aim 2. Validate the disease and cohort specificity of SLE autoantigen signatures. The ~700 million autoantigen data points generated in Aim 1 will be extensively analyzed to identify cohort–specific autoantigens that correlate with ANA staining patterns, ANA reactivities, and patient disease features. The disease- and cohort-relevant autoantigens will be analyzed and categorized in the context of an expanding autoantibody signature database that currently includes 136 sera from patients with four autoimmune diseases and healthy control sera.

Accomplishments under these goals:

Specific Aim 1: Comprehensively identify the molecular targets of ANA standard and SLE autoantibodies.

Major Task 1. Library selections and sequencing for ANA standard (n=17) and SLE patient (n=70) serum samples.

Subtask 1: Autoantigen immunoselection and Illumina NextSeq 500 deep sequencing of n=17 ANA standard sera.

Subtask 2: Autoantigen immunoselection and deep sequencing of n=12 ANA-negative SLE-patient sera and n=12 ANA-positive SLE-patient sera obtained from PSG.

Subtask 3: Select at least 17 additional ANA-negative and ANA-positive SLE-patient sera from Dr. Elizabet Svenungsson's biorepository from the Karolinska Institute.

Subtask 4: Autoantigen immunoselection and deep sequencing of additional ANA-negative and ANA-positive SLE-patient sera (n=29).

Accomplishments:
Subtask 1: This subtask was completed. In 2020, a new cell source was added to our phage library: peripheral blood mononuclear cells (PBMCs) that were infected with Epstein Bar Virus (EBV). All samples described in this report were assayed with a combined phage library of the three cell types discussed in our proposal (astrocytes, white brain matter, and HEp-2 cells) plus the new PBMC + EBV library in order to capture the most diversity. Since characterizing the ANA standards is a foundation for identifying ANA autoantibody diversity, each of the 17 ANA standard sera was assayed and sequenced in three independent experiments to assess variability between assay runs and to ensure precise results, the results of which are described in Task 3.

Subtask 2: ANA-negative SLE-patient sera (n=12) and ANA-positive SLE-patient sera (n=12) were purchased and received from PSG by the end of 2019. In early 2020, these sera were aliquoted and prepared for screening with the Antigenome Platform and all 24 sera were run through the assay and submitted for deep sequencing. Data have been analyzed as described in Task 3.

Subtasks 3-4: In early 2020, Dr. Pisetsky obtained de-identified sera from Dr. Svenungsson's biorepository for these studies. Dr. Svenungsson generously provided 211 sera along with an extensive and comprehensive database describing each patient, their treatments, disease features, and a diverse array of their serological test results. To take advantage of this wealth of patient information, we selected 76 additional patients representing the spectrum of disease, totaling 116 patients' sera rather than the originally proposed 70 samples. While we originally planned to select an additional 17 ANA-negative SLE-patients from Dr. Svenungsson's biorepository, only ANA-positive patients were available.

Major Task 2: ANA assays of SLE sera.

Subtask 1: Screen SLE patient sera to determine/validate ANA reactivity across five clinical ANA assay kits.

Accomplishments:
The samples obtained from PSG and from Dr. Svenungsson's were accompanied by extensive serological analysis data, which were sufficient for our analyses. Therefore, it was unnecessary to test the ANA staining patterns in our lab.

Major Task 3: Bioinformatic processing of all Aim 1 deep sequencing data.

Subtask 1: Process library selection data and identify antibody-reactive antigens that map to the human genome.

Accomplishments: This subtask is complete.

After improvements were made to the phage library, the phage library was sequenced to determine which proteins were available for immunoselection. The four (barcoded) libraries (derived from astrocytes, white brain matter, HEp-2 cells, or PBMCs + EBV) were sequenced with Next Generation Sequencing in two full Illumina HiSeq lane runs, which allowed us to attribute all gene fragments to their derived cell/tissue source. There are at least 17,143 genes represented in the four pooled protein expression libraries used for these assays. The astrocyte library contains protein fragments representing ≥13,219 known human proteins, the HEp-2 library ≥12,406 proteins, the PBMC library ≥14,925 proteins, and the white brain matter library ≥9,561 proteins.

In addition to deep sequencing each library individually, the input library sample (comprised of the four cell source libraries pooled equally) is sequenced as a quality control step in every assay. Therefore, each time we run an Antigenome Platform assay, we add the new control library data to the library gene database; occasionally, genes that were not observed in our initial library sequencing are revealed with this additional sequencing measure. Also, as we run the assays with sera, there are cases where we find new sequences/genes that were not seen in the library sequencing but are revealed when selected and enriched by antibody selection. We also add these data to the library database to obtain the most complete record of expressed genes.

The 116 SLE samples described in Major Task 1 have been assayed and sequenced in full and the resulting sequencing data has been processed through our Antigenome Platform bioinformatics pipeline. In addition, 44 healthy control samples were assayed and sequenced with the improved library for appropriate comparison.

Specific Aim 2: Validate the disease and cohort specificity of SLE autoantigen signatures.

Major Task 4: Generate and analyze autoantigen signatures for the ANA standard sera.

Subtask 1: Perform bioinformatic analyses of library selection results to identify ANA standard sera-reactive antigens and generate autoantigen signatures. Proposed Milestone(s) were to initiate a co-authored manuscript on the autoantigen targets identified by ANA standard sera.

Accomplishments:
This Subtask is complete. The Antigenome Platform data generated in Major Task 1/Subtask 1 (Autoantigen immunoselection and Illumina NextSeq 500 deep sequencing of n=17 ANA standard sera) has been analyzed. Each of the 17 ANA standard sera were assayed three times, which allowed for assessment of reproducibility in

the assay. We found there is 90-96% agreement in autoantigen selection with the same serum sample in three independent assays, demonstrating low inter-assay variability and validating reliable comparisons across experiments.

In analyzing the ANA standard sera, we found that autoantigen targets previously identified by other investigators were also identified by the Antigenome Platform in many cases. Additionally, novel targets were identified in each ANA standard serum sample. While 1-50 antigens have been previously described for each ANA standard serum sample, the Antigenome Platform detected autoantibody reactivity to ~300 antigens per serum sample.

A co-authored manuscript describing the newly discovered autoantigen targets has been initiated and is expected to be submitted in 2023.

Major Task 5. Generate and analyze autoantigen signatures for all patient sera
Subtask 1: Complete selection, assembly, organization, and digitization of all sera patient data.
Subtask 2: Identify and analyze autoantigen signatures that uniquely characterize specific patient subsets.
Subtask 3: Statistical analysis to determine autoantigens targeted exclusively by patient subsets.
Subtask 4: Select additional SLE patients' samples for final round of immunoselection based on the above results.
Subtask 5: Screen all additional/remaining SLE patient sera (n=15-29) for consistency of ANA reactivity across clinical ANA assay kits.
Subtask 6: Bioinformatic interrogation of all identified autoantigens at the domain/exon-level.
Milestone(s): Initiate co-author manuscript on the autoantigen biomarkers that characterize ANA-negative SLE patients and other identified patient subsets.

Accomplishments:
Subtask 1: This subtask has been completed. The Antigenome Platform sequencing data for each patient has been compiled into one database and paired with the patients' respective clinical data.

Subtask 2-4: This subtask has been completed. Dr. Svenungsson published a study in January 2022 (ACR Open Rheumatology) describing four lupus subgroups based on serum selection of known autoantibody targets. These subgroups were used as a foundation for our analysis, wherein we further defined these four subgroups with novel autoantibody profiles. We discovered 32 additional autoantigens defining subgroup 1, 16 for subgroup 2, 6 for subgroup 3, and 14 for subgroup 4 (subgroup 4 was previously characterized by a lack of autoantibody response to known autoantigens); each of these autoantigens are significantly associated with the subgroup compared to the other subgroups (p value $\leq$ .05, q value $\leq$ .1). Bioinformatic analysis shows that the autoantigens selected within each subgroup share motif similarities, which may indicate that antibody cross-reactivity plays a role in the pathogenesis of each subgroup.

Most of the ANA-positive lupus sera have a speckled or homogenous pattern. These data were utilized in combination with the ANA standard sera representing the speckled and homogenous patterns to identify autoantigens associated with these patterns compared to each other and to the ANA-negative patients. We found 11 novel autoantigens significantly associated with the homogenous pattern and selected by the homogenous ANA standard sera (p value $\leq$ .05). We found 4 autoantigens significantly associated with the speckled pattern and present in the speckled ANA standard (p value $\leq$ .05). There are currently no autoantibody biomarkers for ANA-negative SLE patients; however, we identified 16 autoantigens significantly associated with ANA-negative sera (p value $\leq$ .05) compared to homogenous- or speckled-patterned ANA-positive SLE patients.

Subtask 5: The samples obtained from PSG and from Dr. Svenungsson's were accompanied by extensive serological analysis data, which were sufficient for our analyses. Therefore, it was unnecessary to test the ANA staining patterns in our lab.

Subtask 6: This subtask has been completed. Each sequencing run was bioinformatically processed at the domain level, and a database was assembled with domain-level data for each SLE sample.

Milestone(s): A co-author manuscript has been initiated describing the findings in lupus subsets.

Opportunities for training and professional development:
A technician, Joshua Heuler, was hired to the lab in 2021. Mr. Heuler helped with assaying SLE serum samples and preparing them for sequencing. Europe Doan grew in teaching abilities through training Mr. Heuler to run the Antigenome Platform and explaining the project concepts.

How were results disseminated to communities of interest:
To disseminate information on the Antigenome Platform, we presented an abstract for ECTRIMS21, which was held online between 13-15 October 2021. "Genome-wide mapping of patient autoantibody targets to understand and predict multiple sclerosis pathogenesis and patient responses to interferon beta-1a therapy". Europe B. DiCillo[1], Evgueni Kountikov[1], Minghua Zhu[1], Weiguo Zhang[1], Brooke Hayward[2], Danielle E. Harlow[2], Stefan Lanker[2], Jeffrey L. Bennett[3], Thomas F. Tedder[1]; [1]Department of Immunology, Duke University Medical Center, Durham, NC, USA; [2]Neurology & Immunology, US Medical Affairs, EMD Serono, Inc., Rockland, MA, USA, an affiliate of Merck KGaA, Darmstadt, Germany; [3]Departments of Neurology and Ophthalmology, Programs in Neurosciences and Immunology – University of Colorado Anschutz Medical Campus, Aurora, CO, USA.

4. Impact: Nothing to Report yet.

5. Changes/Problems

Changes in approach and reasons for change:
We increased the number of sera to be examined from 70 to 116 to take better advantage of the unique biorepository and rich patient database shared with us by Dr. Svenungsson and to build on Dr. Svenungsson's recently reported findings regarding lupus subgroups.

It became unnecessary to conduct ANA testing on the acquired lupus serum samples in our lab, since the acquired samples had already been ANA-tested, thus the subtasks describing ANA testing were not completed.

Actual problems and delays, and actions to resolve them:
With the COVID-19 pandemic, there were multiple problems that we overcame in 2020 as described in the first annual report.

<u>January 2021-March 2021: Eradicating Lytic Phage Contamination</u>
• After the cDNA inserts from the four cell populations were cloned into the newly constructed Antigenome phagemid, the phagemid libraries were transformed into bacteria for phage production. Despite high transformation efficiency seen by LB-plated bacteria, bacteria grown in liquid culture post-transformation tended to die (this death happened even in the absence of antibiotics).

• After months of investigation, bacterial death was ultimately attributed to a lytic phage contaminating the bacterial cultures. After this infection was uncovered, special care was taken to completely disinfect the lab and stringent cleaning procedures were implemented to prevent re-introduction of the lytic phage into the working environment.
• With the lytic phage eradicated, the phagemid libraries were successfully transformed into bacteria and grown in liquid culture without issue.

April 2021-July 2021: Producing large quantities of phage library

• Given the many changes that had been implemented in the previous two years, the selection of libraries using phage was re-optimized, with the goal of reducing the amount of phage needed and to potentially reduce background in the final sequencing data. Fortunately, the amount of phage needed in each assay was significantly reduced, thereby reducing the months required for optimal phage production from all four libraries (HEp-2, white brain matter, astrocyte, and PBMC + EBV). Stocks of each library were generated and library stocks from each cell source (25% each) were pooled to generate the final library used in assays.
• To keep the starting phage libraries consistent throughout this project, enough phage was produced for all Antigenome Platform assays needed to complete the Aims of two grants supporting these studies, as well as extra phage in case additional assays are required.

August 2021: Next Generation Sequencing of Antigenome Libraries

• The complete Antigenome HEp-2, Brain, Astrocyte, and PBMC phage libraries were individually sequenced to quantify the diversity of antigens represented among all the libraries. NGS revealed that >17,000 genes are represented across all four libraries. Unfortunately, the Brain and HEp-2 libraires were sequenced at low depth (~50 million reads), so deeper sequencing of these libraires to achieve 100 million reads is underway. It is hypothesized that, after additional sequencing of these libraries, the number of genes represented in the combined libraries will be closer to the target goal of 20,000, representing the entire human genome.

Changes that had a significant impact on expenditures:
The reductions in personnel during the COVID pandemic have reduced personnel expenditures as well as supply expenditures. These funds have been used to complete the originally proposed studies during the period of the No-cost-extension.

Changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents: Nothing to Report.

6. Products: Nothing to Report yet.

7. Participants & Other Collaborating Organizations

Individuals that have worked on the project:

Name:	Thomas F. Tedder
Project Role:	PI
Research Identifier: ORCID	0000-0003-0653-7354
Person month worked:	1.20 CM

Contribution to Project:	Lead, directed and supervised the research effort
Funding Support:	

Name:	David S. Pisetsky
Project Role:	Co-investigator
Research Identifier: eRA C	PISETSKY
Person month worked:	0.60 CM
Contribution to Project:	Reviewed experimental plans, serum collection, data analysis
Funding Support:	

Name:	Cliburn C. Chan
Project Role:	Collaborator
Research Identifier: eRA C	CCCHAN
Person month worked:	0.24 CM
Contribution to Project:	Biostatistical analysis and bioinformatics
Funding Support:	

Name:	Europe B. Doan
Project Role:	Graduate Student
Research Identifier: eRA C	0000-0001-5874-9514
Person month worked:	6.00 CM
Contribution to Project:	laboratory research, data analysis, bioinformatics analysis
Funding Support:	

Changes in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period:

None to report.

Other organizations involved as partners: Nothing to Report

8. **Special Reporting Requirements**: None

9. **Appendices**: None