ImmunoSignature™ Technology Differentiates Patients with Systemic Sclerosis and Internal Organ Involvement

Authors: Lorinda Chung[1], Robert Gerwien[3], Kathryn Sykes[3], Kathleen Jia[1], Joseph Barten Legutki[3], Antonia Valenzuela Vergara[1], Jon Melnick[3], Lisa Zhu[1], Theodore M. Tarasow[3], David Fiorentino[2]

Corporate Office
2420 Camino Ramon
Ste. 125
San Ramon, CA 94583

Chandler Innovations Center
145 South 79th Street
Chandler, AZ 85226

Abstract # 59897

Disease and Technology Background

Systemic Sclerosis, or Scleroderma (SSc):

- Diagnosis of SSc is difficult due to the complexity of manifestations and overlap with other autoimmune diseases (AID).
  - SSc-specific autoantibodies have been identified[1] and are included in the updated classification criteria (2013 ACR/EULAR), but disease manifestations are often advanced by the time patients fulfill these criteria.
  - Heterogeneity of clinical presentation, range of internal organ involvement, and differences in rates of disease progression make counseling and management of each individual patient's disease challenging.
- The goals of this study were:
  - To identify reproducible antibody signatures that differentiate SSc patients from healthy individuals
  - To explore signatures that distinguish SSc patients from those with other AIDs such as dermatomyositis (DM).
  - To explore signatures that distinguish SSc patients with or without specific internal organ complications such as gastric antral vascular ectasia (GAVE).

ImmunoSignature™ Technology:

- The IMS technology uses arrays of hundreds of thousands of unique peptides designed from all possible strings of amino acid rather than proteome-derived designs. This enables broad surveys of an individual's antibody binding repertoire from a small sample of blood, serum, or plasma[2].
- This Technology is a diagnostics and disease monitoring platform that has shown utility in a variety of immune-mediated diseases[3,4].
- Here we explore its applicability to SSc compared with DM and a variety of other AIDs.

Characteristics of the Technology:

- Comprehensive use of antibodies as whole body surveillance and alert molecules.
- Design of peptide libraries that maximize the diversity of available binding sites, rather than target specific sequences found in nature.
- Synthesis of the peptide libraries peptides (~126,000) upon silicon wafers so as to optimally present the mimetic-ligands to blood-antibodies.
- Use of addressable HealthTell IMS arrays to quantitatively measure the competitive binding events.
- Analysis of antibody binding patterns provides informative health readouts.
- Mimetic-peptide analysis can indicate the cognate epitope targets.

Array & Assay Qualifications:

- The HealthTell IMS arrays include control sequences that match epitopes of well characterized monoclonal antibodies (mAbs).
- Binding patterns to control and library peptides were measured to qualify the array manufacturing process and the IMS assay process.
- Inter wafer signal precision was determined by testing plasma sample replicates on arrays from different wafers and calculating the coefficients of variation (CV) for all library peptides[5].

Methods and Study Design

Assay Process:

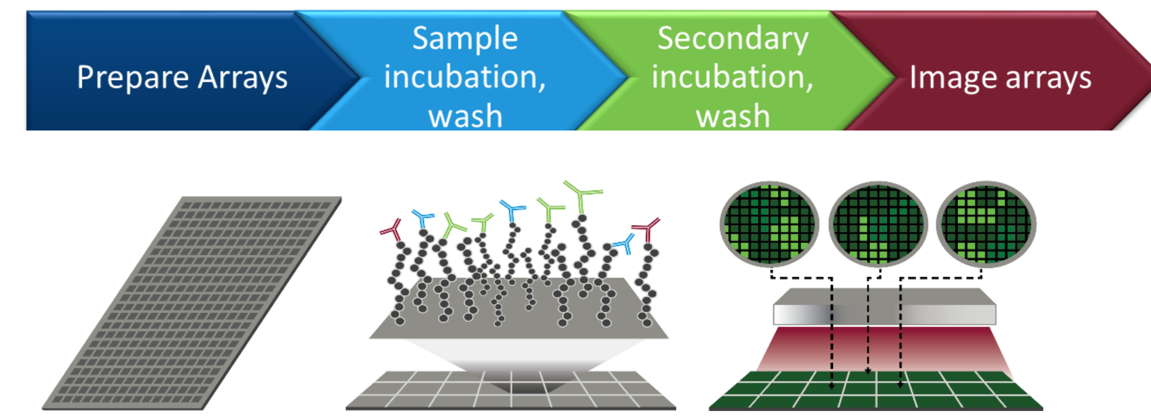

FIGURE 1: Assay workflow. A small blood sample is diluted and incubated on the array. After a wash step, peptide-bound antibodies are fluorescently tagged. Binding data is acquired using a fluorescent imager. Peptide-binding patterns associated with specific disease state are identified.

- Plasma samples from a study population of 719 unique donors were evaluated; the cohort was comprised predominantly of SSc (n=301), DM (n=205) and healthy individuals (n=118). Other AIDs (n=95) included MCTD, UCTD, systemic lupus erythematosus, polymyositis, and localized scleroderma (morphea). Age and gender distributions were similar in all groups.
- All met ACR classification criteria for each specific disease at diagnosis.
- A panel of 84 control samples was used to facilitate assay qualifications.
- The IMS assay was used to detect array-bound antibodies. Features most discriminating the SSc contrasts were identified using a Welch's t-test.

TABLE 1: Clinical and Demographic Characteristics of Systemic Sclerosis Cohort

Description	Representation
Female (% total)	263 (87%)
Age (Mean range)	54 (20-84)
Disease Duration (Mean Years ± SD)	8.4 +/- 9.08
Diffuse (%)	114 (38.8%)
SCL-70+ (%)	58 (21.7%)
RNA pol III+ (%)	37 (30.1%)
ACA+ (%)	91 (37.0%)
ILD+ (%)	91 (42.5%)
PAH+ (%)	36 (51.4%)
SRC+ (%)	11 (3.7%)
GAVE+ (%)	13 (7.2%)

- Classifiers were developed on the list of 25 to 10,000 significantly discriminating features using Support vector machines (SVM). SVMs are supervised machine learning algorithms used for classification of each contrast.
- A 5-fold cross-validation (CV) analysis was used to assess performance. This divides the data randomly into 5ths and trains an algorithm on 4/5ths , then tests it on the left out 1/5th. This is repeated 4 more times so that all samples are included in the test set.
- Feature selection bias was controlled by including feature selection within the CV loop. Sampling bias was controlled by repeating the CV process 100 times.

SSc ImmunoSignature™ Results

Classification Performance:

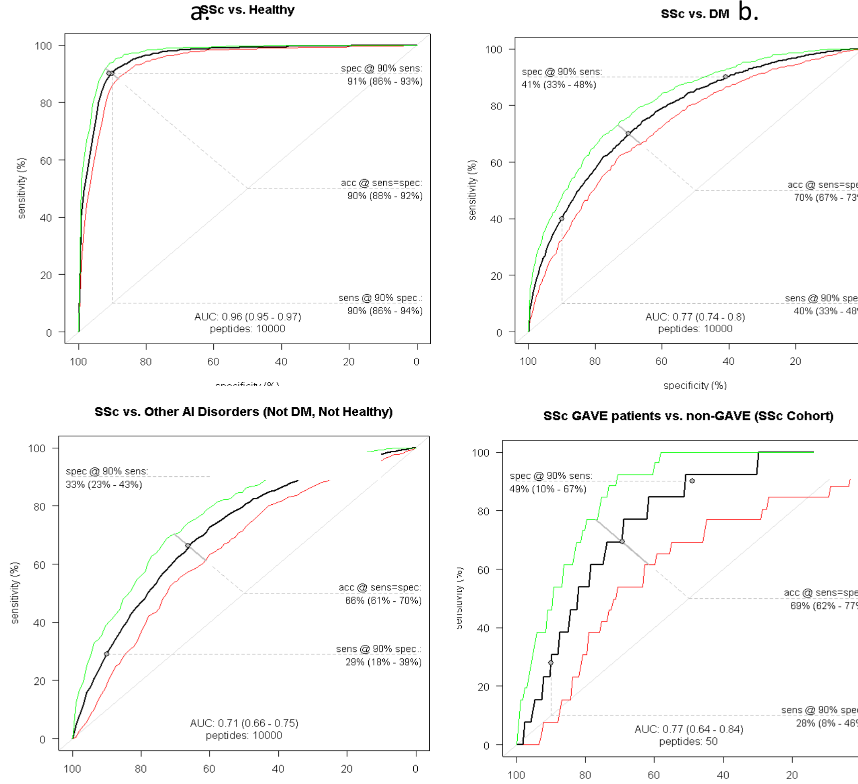

FIGURE 3: Receiver operating characteristic (ROC) curves describe the cross-validated performance of IMS for AIDs. For example (a) scleroderma disease discrimination versus healthy, n=301/118, (b) SSc vs. dermatomyositis, n=301/205, or (c) SSC vs. a heterogeneous cohort of other autoimmune diseases, n=301/95. IMS performance is also shown in identifying: (d) SSc patients whose disease is complicated by organ involvement (GAVE) vs. those whose is not, n=13/166. Model sizes (k) from 50 to 10,000 peptides whose antibody-binding characteristics differentiated health status groups were explored as inputs to these SVM classifiers. The optimal model size was used to build the ROC curves.

TABLE 2: Classification Performance Estimates of IMS for SSc Diagnosis and Prognosis contrasts

Contrast	AUC	Sens. @ 90% Spec.	Spec. @ 90% Sens.	Accuracy @ Sens. = Spec.
SSc vs Healthy	0.96 (0.95-0.97)	90% (86-94%)	91% (86-93%)	90% (88-92%)
SSc vs DM	0.77 (0.74- 0.8)	40% (33-48%)	41% (33-48%)	70% (67-73%)
SSc vs Other AID	0.71 (0.66-0.75)	29% (18-39%)	33% (23-43%)	66% (61-70%)
GAVE+ vs GAVE-	0.77 (0.64-0.84)	28% (8-46%)	49% (10-67%)	69% (62-77%)
ILD+ vs ILD-	0.68 (0.64-0.72)	23% (13-33%)	31% (21-41%)	63% (59-68%)
Renal Crisis+ vs Crisis-	0.72 (0.60-0.82)	27% (3-53%)	42% (12-62%)	65% (55-76%)
PAH+ vs. PAH-	0.44 (0.36-0.54)	8% (1-20%)	5% (0-12%)	47% (37-56%)

Sequence Analysis of Classifying Library Peptides :

Library peptides were aligned to human proteome RefSeq release 84, corresponding to human genome build GrCh386 using the longest transcript variant for each unique gene ID. The alignment algorithm used a modified BLAST strategy7, requiring a seed of 3 amino acids (aa), a gap penalty of 4 aa, and a scoring matrix of BLOSUM628 modified to reflect the aa composition of the array9 . These modifications increase the score of similar substitutions, remove penalties for aa's absent from the array and score all exact matches equally.

To generate an alignment score to a protein for a set of classifying library peptides, those that yield a positive BLAST score are assembled into a matrix, with each row of the matrix corresponding to an aligned peptide and each column corresponding to one of the consecutive aa's that comprise the protein. Gaps and deletions are permitted within the peptide rows for alignment to the protein. In this way, each position in the matrix receives a score associated with the paired aa of the peptide and protein. Each column, corresponding to an aa in the protein, is then summed to create an overlap score; this represents coverage of that aa position by the classifying peptides. To correct this score for library composition, another overlap score is calculated using an identical method for a list of all array peptides. This allows for the calculation of a peptide overlap difference score, s, at each aa position via the equation s=a-(b/d)*c . In the equation, a is the overlap score from the classifying peptides, b is the number of classifying peptides, c is the overlap score for the full library of peptides and d is the number of peptides in the library.

To convert these scores at the aa level (s) to a full-protein statistic, the sum of scores for every possible tiling 20-mer epitope within a protein is calculated. The final protein epitope score, S, is the maximum along this rolling window of 20 for each protein. A similar set of scores is calculated for 100 rounds of peptides, equal to the number of classifying peptides, iteratively selected from the library. This provides a p-value for each S. POL2RL has a p-value =0.00005

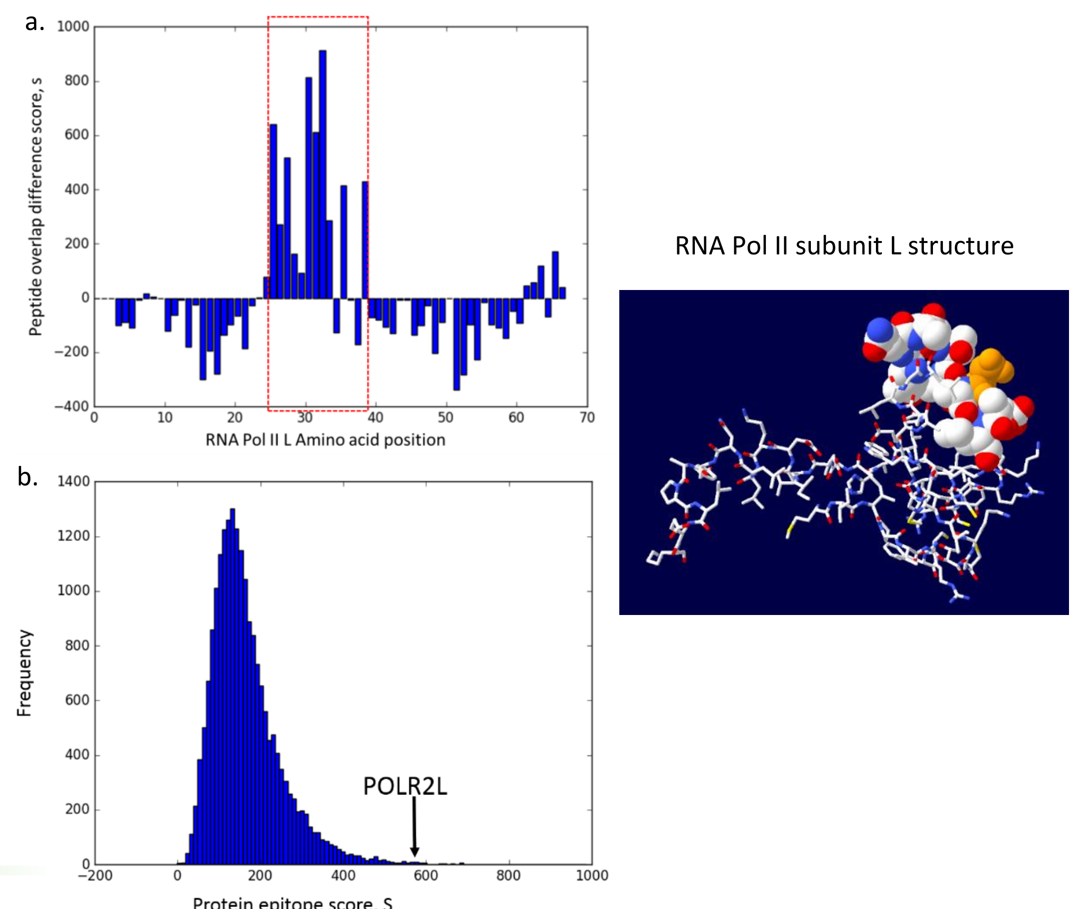

FIGURE 4: Library peptides that distinguish health states can be analyzed relative to the human proteome to indicate likely immunogenic autoantigen targets. An exemplary library peptide hit is presented. (a) The peptide overlap difference scores, s calculated for the alignments of IMS peptide-motifs are plotted alongside the RNA Pol II subunit L aa positions. Peptides from the SSc vs. healthy contrast showed significant alignment with RNA pol II subunit L, ranking 35 of 20,378 human proteins. The ball and stick model to the right shows the structure of RNA pol II subunit L. The region displayed in balls corresponds to the aa positions marked with a red box within the graph. The highest scoring aa is aspartic acid, D, in the center of the RNA pol cluster, is shown in orange. (b) Histogram displaying the distribution of protein epitope scores, S, for each protein in the human proteome mapped against the SSc vs healthy classifying peptides. POLR2L's score is 583.

Summary

Results overview:

- A strong classifier distinguished SSc patients from healthy individuals.
- Additional classifiers differentiated SSc from DM or from other AIDs.
- Preliminary classifiers were built to distinguish SSc patients whose disease was complicated by gastric antral vascular ectasia (GAVE).
- The IMS library peptides that significantly distinguish SSc from healthy individuals carry motifs that align with a set of human proteins. Several of these are AID-related autoantigens.
- RNA pol II L was in the top 1% of hits. It has been shown to be immunogenic in SSC patients. The well-characterized RNA pol III subunit A antigen was not indicated; however, only 30% of the patients in the cohort were positive for anti-RPC155 antibodies.

Conclusions:

- Reproducible binding patterns produced by peripheral-blood antibody repertoires on HealthTell's mimetic-peptide arrays can differentially identify SSc patients from healthy individuals and from those with other AIDs.
- A distinctive signature was established for SSc patients whose disease was complicated by GAVE, a serious disease manifestation.
- The ImmunoSignature technology may be useful in the development of new diagnostic tests for SSc and related autoimmune diseases.

References:

1. Van den Hoogen et al. (2013). Arthritis Rheum. 65(11):2737-47
2. Sykes et al. (2013) Trends in Biotech. 31: 45-51.
3. Legutki et al.. (2014) Nat. Commun 5: 4785.
4. Stafford et al. (2014) PNAS 11: E3072-3080.
5. Rowe et al. PLOSNTD, manuscript in review
6. https://www.ncbi.nlm.nih.gov/refseq/
7. Altschul, S.F. & Gish, W. (1996) Meth. Enzymol. 266: 460-480
8. Henikoff, J.G. (1992) Proc. Natl. Acad. Sci. USA 89: 10915–10919.
9. States, D.J., Gish, W., Altschul, S.F. (1991). Methods 3: 66-70.

Acknowledgements:

- The authors would like to acknowledge and thank David Lomeli and Kathryn Nielsen for their work in conducting the assays; Scott Melville and Glenn Hein for designing sample layouts described in this study; and Patrick Walsh for directing the peptide array syntheses.

Author Affiliations:

[1]Stanford University Medical Center
[2]Stanford University School of Medicine
[3]HealthTell, Inc.