Pilot study: prognostic biomarkers for interstitial lung disease in systemic sclerosis

Mantero, Julio C.
BOSTON UNIVERSITY
SCHOOL OF MEDICINE

Thesis

PILOT STUDY: PROGNOSTIC BIOMARKERS FOR INTERSTITIAL LUNG DISEASE IN SYSTEMIC SCLEROSIS

by

JULIO C. MANTERO
B.S., Penn State University, 2013

Submitted in partial fulfillment of the requirements for the degree of Master of Science 2016
Approved by

First Reader

Robert Lafyatis, M.D.
Professor of Medicine

Second Reader

Romy Christmann, M.D. Ph.D.
Associate Professor of Medicine

Third Reader

Robert Simms, M.D.
Professor of Medicine
Section Chief, Rheumatology
ACKNOWLEDGMENTS

I want to thank my thesis readers Dr. Christmann and Dr. Simms for their guidance, advice and wisdom during this process. I would also like to thank my mentor Dr. Lafyatis for giving me the opportunity to work on this project. Without his valuable guidance and knowledge this work would not have been completed. I would also like to thank the people in the Clinical Investigation program for their helpfulness and patience since the beginning of my thesis to the very end. Lastly, I would also like to thank my lab colleagues for their suggestions and support during this process.
PILOT STUDY: PROGNOSTIC BIOMARKERS FOR INTERSTITIAL LUNG DISEASE IN SYSTEMIC SCLEROSIS

JULIO C. MANTERO

ABSTRACT

Interstitial lung disease is one of the main causes of mortality in Systemic Sclerosis. The course of the disease is clinically variable where patients can suffer from a range of stable disease to rapid progressive clinical deterioration. Therefore, it is important to identify biomarkers that can predict the clinical course of patients in order to provide early treatment. We evaluated 1129 proteins utilizing novel high-throughput SOMAlogic proteomic technology from the serum of 13 ISSc, 13 progressive ILD and 11 stable ILD patients. Calpain-1 was significantly elevated in progressive ILD patients (median 15129 RFU, 11091-24561) compared to ISSc patients (12759, 9904-15498, p=0.0015) and stable ILD patients (11876, 10271-14249, p=0.0005). Coagulation Factor V was significantly lower in the progressive ILD patients (7161 RFU, 2140-8296) compared to ISSc patients (10311 RFU, 6396-12260, p=0.001) and stable ILD patients (9646 RFU, 6510-11941, p=0.0016). The combination of Coagulation Factor V and Calpain-1 produced an area under the curve of 0.97 (95% CI, 0.921-0.99), sensitivity of 99% and specificity of 91% for the identification of progressive ILD. We have identified a combination of proteins that show potential to be prognostic biomarkers for ILD in SSc.
# TABLE OF CONTENTS

TITLE ............................................................................................................... i

COPYRIGHT PAGE ......................................................................................... ii

READER APPROVAL PAGE .............................................................................. iii

ABSTRACT ........................................................................................................... v

TABLE OF CONTENTS ....................................................................................... vi

LIST OF TABLES ................................................................................................. viii

LIST OF FIGURES ............................................................................................. ix

LIST OF ABBREVIATIONS .................................................................................. x

INTRODUCTION .................................................................................................. 1

Systemic Sclerosis ............................................................................................. 1

Interstitial Lung Disease .................................................................................... 2

Treatment Options ............................................................................................. 3

Biomarkers ........................................................................................................ 5

Objective/Aims .................................................................................................. 8

METHODS ......................................................................................................... 9

Ethics Statement ............................................................................................... 9

Study Subjects .................................................................................................. 9

Sample Collection ............................................................................................ 11
LIST OF TABLES

Table | Title                                           | Page |
----- |-------------------------------------------------|------|
1     | Study Subjects Characteristics                   | 14   |
2     | Comparison of Serum Biomarkers                   | 19   |
3     | Logistic Regression Model ROC measurements       | 21   |
4     | Combined Logistic Regression Model ROC measurements | 22   |
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A</td>
<td>FVC trajectories for progressive ILD subjects</td>
<td>10</td>
</tr>
<tr>
<td>1B</td>
<td>FVC trajectories for stable ILD subjects</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>Heatmap</td>
<td>16</td>
</tr>
<tr>
<td>3A</td>
<td>Comparison of serum Coagulation Factor V in lSSc, stable ILD (STA) and progressive ILD (PRO)</td>
<td>17</td>
</tr>
<tr>
<td>3B</td>
<td>Comparison of serum Calpain-1 in lSSc, stable ILD (STA) and progressive ILD (PRO)</td>
<td>18</td>
</tr>
<tr>
<td>3C</td>
<td>Comparison of serum PARC in lSSc, stable ILD (STA) and progressive ILD (PRO)</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>Analytes ROC curves</td>
<td>22</td>
</tr>
<tr>
<td>5</td>
<td>Combined analytes ROC curves</td>
<td>22</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

AUC .................................................................................................................. Area under the curve
ILD ..................................................................................................................... Interstitial Lung Disease
PRO ..................................................................................................................... Progressive ILD
ROC ................................................................................................................... Receiver operator curves
STA ..................................................................................................................... Stable ILD
SSc ..................................................................................................................... Systemic Sclerosis
INTRODUCTION

Systemic Sclerosis

Systemic sclerosis (SSc) or scleroderma is a rare, autoimmune, multisystem connective tissue disorder characterized by vascular damage, autoimmunity and fibrosis of the skin and other internal organs.\(^1\)\(^{-}\)\(^3\) The incidence of SSc in the United States is estimated to be about 20 cases per 1 million adults, while its prevalence is estimated to be about 276 cases per million adults.\(^4\) The disease predominantly affects women, with an approximate 5:1 ratio of women to men diagnosed.\(^4\) The disease seems to be most prevalent amongst middle-aged women (35-50 years old) but it can affect women of any age as well as men and children.\(^4\) The median time of survival of patients with the disease is of approximately 11 years after diagnosis.\(^4\) The pathogenesis of SSc is not well understood. However, it is known that there is a relationship between inflammatory, vascular, and fibroblast dysfunction. The clinical manifestations of SSc are varied which complicates early and accurate diagnosis of potentially progressive patients.

The disease is usually classified into two major clinical subgroups based on the extent of skin involvement; limited cutaneous systemic sclerosis (lcSSc) and diffuse cutaneous systemic sclerosis (dcSSc).\(^5\)\(^,\)\(^6\) Patients classified with dcSSc have more widespread skin fibrosis involving the trunk and proximal extremities of the body, while patients with lcSSc have less extensive skin involvement usually limited to the distal extremities including hands, feet, forearms and face.\(^6\) To evaluate the extent of the skin involvement, physicians utilize the Modified Rodnan Skin Score (MRSS). The score is
obtained during an initial physical examination where the physician manually palpates 17 areas of the body: fingers, hands, forearms, arms, feet, legs, thighs (in pairs) and face and gives each site a score (0-3). The total skin score is the sum of the individual sites, with MRSS ranging from zero to fifty-one. Higher MRSS describes more extensive disease.

**Interstitial Lung Disease**

At present, pulmonary disease is the main cause of morbidity and mortality in patients with interstitial lung disease (ILD) leading to pulmonary fibrosis (PF), and vascular obliteration leading to pulmonary arterial hypertension (PAH).\(^1,7\) Together pulmonary manifestations account for 60% of SSc related deaths.\(^1,8\) The prevalence of ILD in SSc patients ranges from 75% to 90%. Amongst the SSc subgroups, ILD seems to be more prevalent in patients with dcSSc, while PAH appears to be more common in patients with lcSSc.\(^1\) The development of ILD often manifests during the first 3 years of the disease, however some studies have suggested that it can manifest even 7 years from the onset of SSc.\(^9\) SSc patients with ILD tend to have a higher mortality than patients without organ involvement. Patients with lung involvement have a nine-year survival of 30% while patients without lung involvement have a nine-year survival of 72%.\(^10\)

The presence of antinuclear antibodies (ANA’s) is common in SSc patients; present in nearly 90% of patients. In the context of ILD, anti-Scl-70 (anti-topoisomerase I) antibodies are associated with a higher prevalence of pulmonary fibrosis and are found in approximately 40% of patients with dcSSc and in less than 10% of patients with lcSSc.\(^11,12\) On the other hand, anticentromere antibodies are less frequently found in
patients with lung disease and are associated with a lower prevalence of pulmonary fibrosis.12

In the clinical setting patients suffering from ILD may complain of shortness of breath, dyspnea, dry cough, atypical chest pain and general weakness.1,7 High resolution computed tomography (HRCT) is commonly utilized to identify interstitial abnormalities in the lungs of SSc patients, and is the standard tool for diagnosis of ILD in SSc. Septal and subpleural line opacities, ground glass opacities and subpleural cysts are early signs of ILD. Progression of ILD is characterized by the formation of honeycombing and bronchiectasis.13,14 Formation of fibrotic tissue in the lungs results in restrictive changes in the lung function of patients.13,14 Pulmonary function tests (PFT’s) are performed to identify and monitor the extent and progression of these changes in the lungs. Force vital capacity (FVC), defined as the amount of air that can be forcibly exhaled from the lungs after taking the deepest breath possible, is the most commonly monitored value by physicians to determine the severity and progression of the disease. At baseline, the severity of ILD can be classified on a scale where predicted FVC values of >80% are considered normal, 70-79 % (mild), 50-69% (moderate) and 50% (severe).7 Reductions of >10% of FVC from baseline over a period of 3-12 months indicates disease progression.7

Treatment Options

Current treatments for ILD in SSc include the use of corticosteroids, immunosuppressant agents, autologous hematopoietic stem cell transplantation and lung transplantation. Corticosteroids such as Prednisone are commonly utilized in SSc but
their use in high doses have been associated with the onset of scleroderma renal crisis (SRC) and are mostly utilized in lower doses in combination with Cyclophosphamide. Cyclophosphamide is the most utilized immunosuppressant for the treatment of ILD in SSc. The effectiveness of Cyclophosphamide in SSc has been extensively studied but variable results regarding its effectiveness, small improvements in pulmonary function alongside the numerous side effects associated with the drug has emphasized the need for alternative less toxic therapies.\textsuperscript{15,16}

Mycophenolate mofetil (MMF) is another immunosuppressive agent that has become an alternate option for patients that are not able to tolerate Cyclophosphamide. In retrospective and prospective studies its use has been associated with improvement or stabilization of lung function however such studies have involved been small samples of subjects, thus requiring larger controlled studies will be needed to truly assess its effectiveness.\textsuperscript{15} Hematopoietic stem cell transplantation (HSCT) is another approach that attempts to downregulate the aberrant immune reaction that appears to drive the disease. Various clinical trials have demonstrated that treatment of severe SSc patients with HSCT can improve MRSS and FVC over baseline measurements.\textsuperscript{17} The caveat with this treatment is that it is associated with a high risk of mortality and toxicity.\textsuperscript{18}

Thus, treatment options for SSc-ILD come with significant treatment-associated morbidity and mortality.\textsuperscript{17} Currently the ability to predict the progression of ILD in SSc patients at time of diagnosis is extremely challenging because the disease is clinically variable and in some cases highly progressive, while in others it stabilizes without the need of treatment.\textsuperscript{19} Therefore, identifying the patient population at higher risk of disease
progression is extremely important. The identification of one or multiple biomarkers that
distinguish patients likely to have progressive ILD would allow physicians to identify
patients who need to be treated with these aggressive treatments and spare those who are
at a lower risk for clinical progression. Currently there are no clinically validated
prognostic biomarkers for ILD that would enabled physicians predict the rate of
progression of ILD over time, indicating the importance of discovering and validating
new biomarkers to address this unmet clinical need.20,21

Biomarkers

Biomarkers are objective outcome measures that are used to measure a biological
process or assess the response to a therapeutic intervention. They provide valuable
information for more efficient performance of clinical trials while simultaneously
creating a bridge to understanding disease pathogenesis.22 Biomarkers already permeate
clinical care as commonly used laboratory tests. Some examples are laboratory tests to
assess TSH levels for thyroid disease, PSA for prostate cancer or CK levels to assess the
occurrence of muscle injury and myocardial infarction.22 There are different types of
biomarkers that can be applied for different purposes including diagnostic, predictive,
prognostic and pharmacodynamics biomarkers. Diagnostic biomarkers are utilized to
identify patients who suffer from a particular disease, condition or complication.
Predictive biomarkers provide information about the likelihood of a patient responding to
a specific treatment. Pharmacodynamic biomarkers are utilized to determine the
effectiveness of a therapeutic intervention and are often utilized as endpoints in clinical
trials. Prognostic biomarkers indicate the future clinical course of the patient in the
context of a particular clinical outcome without regard to treatment.\textsuperscript{23} Identifying prognostic biomarkers for ILD in SSc would directly impact the treatment plan of patients that would suffer from the progressive disease by identifying patients at risk earlier and provide early therapy.\textsuperscript{24}

Currently, the only way to identify subjects whose ILD will progress is by performing multiple PFT’s and assessing whether there are any changes in FVC. This poses a problem because while patients are waiting to get new PFT’s they are not being treated and their lungs are potentially suffering irreversible damage. In addition, clinical trials could also benefit from prognostic biomarkers. Trial populations could be enriched with subjects who suffer from progressive disease making trials designed to test progression of the disease more robust.\textsuperscript{25} Therefore, prognostic biomarkers could help identify those subjects who will develop progressive ILD, reduce the number of subjects in clinical trials, and avoid exposing subjects with stable disease to the risks potentially associated with the investigational therapy.

Other researchers have investigated several potential biomarkers for progressive disease. High levels of CXCL-4 in the plasma have been shown to correlate with the presence of lung fibrosis.\textsuperscript{26} However, this study included a wide spectrum of SSc patients and did not specifically look at whether CXCL-4 levels were different in patients with progressive ILD compared to stable ILD patients and the results have yet to be independently validated. Previous studies have identified several serum proteins like Krebs von Lungen-6 (KL-6), surfactant protein-D (SP-D) and CCL18 as candidate prognostic biomarkers.\textsuperscript{19,22,24,27–29,30} An increase of these markers has been seen in
patients with ILD, however their validation as prognostic biomarkers has not been possible because of inconsistent and variable measurements when correlating with FVC.19

Other groups have studied the bronchoalveolar lavage fluid (BALF) of patients to understand more about the pathogenesis of ILD and to find biomarkers.31,32 One of the studies identified high concentrations of CXCL5, CXCL8 and S100A8/A9 in the BALF of patients that were statistically significant when compared to healthy controls. These analytes also correlated with findings on HRCT and may serve as markers for the presence and extent of lung fibrosis31. High levels of cytokines in BALF were observed in another study where SSc patients with ILD had higher concentrations of IL-4, IL-6, IL-8 and CCL2 compared to controls and correlated negatively with FVC.32 This approach offers the opportunity of studying the lung tissues more closely, however obtaining BALF is more invasive and thus a more difficult source to obtain multiple measurements. Thus, one could question their practicality in a clinical setting.19

In this study we utilize a novel proteomic technology called SOMAscan to assess and identify, quantitative protein differences between subjects with stable ILD and progressive ILD in order to identify potential prognostic biomarkers. This is a high throughput technology that measures a higher number of proteins in serum than commonly used mass-spectrometry. Mass-spectrometry is difficult to use on serum samples due to the loss of sensitivity due to the high abundance from several very common serum proteins such as albumin, making it hard to identify biomarkers.33 The technology is composed of a capture array called SOMAscan using a Slow Off-rate
Modified Aptamer (SOMAmer) that transforms protein signal to nucleotide signal that allows quantification of 1,129 proteins by relative florescence. This technology has been utilized widely utilized in other studies with the purpose of identifying biomarkers in other disease and has provided positive results. Studies include the identification of biomarkers in inflammatory arthritis and identification of plasma protein biomarkers associated with cognitive decline in Alzheimer’s disease patients.\textsuperscript{34,35}

Also, looking at proteins in the serum is a less invasive and more practical way of biomarker identification since serum is easily obtainable.

**Objective/Aims**

The aim of this study is to discover and develop one or multiple prognostic biomarkers that would allow identifying subjects that will develop progressive ILD. Previously several biomarkers for ILD have been studied, however results have not been validated successfully. ILD is one of the main causes of mortality in SSc thus discovering new and better biomarkers is of importance. This study will use a dataset of 1,129 analytes obtained from a total of 40 samples that were sent for proteomic profiling to SOMAlogic. This study is unique because it is the first study to utilize this technology to identify prognostic biomarkers for ILD in SSc. The obtained dataset will be analyzed to identify differences in protein expression between progressive and stable ILD subjects. A shortlist of analytes will be developed, by calculating false discovery rates, to identify those that might serve as potential biomarkers. Then by using linear regression models we will attempt to identify one or a combination of analytes that will identify subjects that will develop progressive ILD.


METHODS

Ethics Statement

This study was conducted under a protocol approved by the Institutional Review Board of Boston University Medical Center. All patients signed informed written consent forms approved by the Boston University Medical Center Institutional Review Board.

Study Subjects

All the samples analyzed in this study consisted of patients enrolled the Scleroderma Clinical Repository (SCaR) located at Boston University Medical Center. The repository contains a population of approximately 500 subjects that have been recruited from the Boston Medical Center Rheumatology and Pulmonary clinics. All study subjects have met the criteria for limited cutaneous systemic sclerosis (lSSc) as defined by the American College of Rheumatology (ACR). Subjects with lSSc were stratified into three groups those with no ILD (16 subjects), those with progressive ILD (13 subjects) and those with stable ILD (11 subjects). LSSc subjects showing evidence of ILD and confirmed by HRCT where considered to have ILD (24 subjects).

ILD subjects were further stratified in two groups those with progressive disease and those with stable disease. Subjects were determined to be progressive or stable based on the total trend in FVC measurements from serial PFT’s that were collected 24 months before and after sample collection, which were obtained through the available database data and medical records review. Each of the subject’s FVC measurements was graphed to determine which subjects had progressive disease versus those who had stable disease.
Subjects with PFT’s that showed an overall trend of decline (>5% decrease) of FVC over a 48-month period (Figure 1A) were classified as progressive (13 subjects). Subjects whose PFT’s showed minimal changes, defined as less than 5% decrease, or increasing FVC over 48 months (Figure 1B) were classified as stable (11 subjects).

Demographic characteristics and clinical manifestations were assessed for the subjects including history (age and sex,), and medical records (disease duration, ILD duration, HRCT, PFT’s, concomitant medications,). LSSc disease duration was measured from the date the subject was first diagnosed with lSSc by a rheumatologist. ILD duration was measured from the date of diagnosis and confirmed by HRCT.

**Figure 1A.** FVC trajectories for progressive ILD subjects
Sample Collection

Biomarkers were assayed from stored serum samples that were collected from patients as part of the SCaR repository at Boston University School of Medicine. Serum was collected in standard Serum Separator Tubes (BD Vacutainer) in accordance with manufacturer’s instructions and mixed and centrifuged at 1,200g at room temperature for 15 minutes. Serum was then aliquoted into collection tubes in volumes of 250ul, 500ul and 1000ul and frozen at -80 Celsius immediately after collection.
**Biomarker measurement**

Proteins were measured using a high throughput technology called ‘SOMAscan’ (SomaLogic, Inc, Boulder, Colorado). This approach uses chemically modified nucleotides to transform a protein signal to a nucleotide signal that can be quantified using relative fluorescence on microarrays. This assay simultaneously measured the level of 1,129 human proteins in a single sample.

**Statistical Analysis**

Continuous data is presented as the median and range and categorical data is presented as percentages. Proteomic data was analyzed utilizing R-(version 0.98.1102) and Prism (Graphpad version 6.0g). Difference in protein expression between progressive and stable ILD was calculated by Wilcoxon rank sum test. Difference in protein expression between ILD vs no ILD was also calculated by Wilcoxon rank sum test. Resulting p-values from the two Wilcoxon rank sum tests were combined by Fischer's test. Then, to reduce Type-I error, false discovery rates (FDRs) was implemented to adjust p-values for multiple comparisons and adjusted p-values of less than 0.1 were considered to indicate potential statistical significance. Fifty analytes had adjusted p-values less than 0.1 and were used for unsupervised hierarchical clustering to assess whether the progressive and stable groups independently separated based solely on association between analytes within the phenotypic groups. Then, seven analytes were selected based on their association with other analytes, each of the seven analytes that were selected was in a different cluster in order to avoid selecting overly similar analytes. The Kruskal-Wallis test was used to assess for differences across the three groups and
correcting with Dunn’s multiple comparison post-test was employed to analyze differences in the mean expression between lSSc with no ILD, stable and progressive groups for these seven analytes. Statistical significance was considered as p-values of <0.05. Disease status (dependent variable) was dichotomized in order to build logistic regression models to determine if selected analytes (independent variable) predict disease progression versus stable. Linear regression is a common statistical method utilized to test whether one or more independent variables determine an outcome. The utility of the biomarkers was assessed by utilizing sensitivity and specificity measurements and the area under the receiver operating characteristic (ROC) curve in order to identify specific cut-off values for the proteins.
RESULTS

Baseline characteristics of the study subjects are shown in Table 1. SSc disease duration in the progressive ILD group was found to be shorter when compared to lSSc (p=0.04) and to the stable group (p=0.02). The baseline FVC, defined as the closest value to sample collection, was found to be significantly lower in both ILD groups (progressive (p=0.0001), stable (p=0.0012)) when compared to the no ILD lSSc group. ILD disease duration was found to be significantly greater (p=0.0024) in the stable ILD group when compared to the progressive ILD group.

<table>
<thead>
<tr>
<th>Subject Demographics</th>
<th>LSSc (16)</th>
<th>PRO (13)</th>
<th>STA (11)</th>
<th>LSSc vs PRO</th>
<th>LSSc vs STA</th>
<th>PRO vs STA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>Median (range)</td>
<td>56.50 (25-70)</td>
<td>47 (31-59)</td>
<td>59 (29-77)</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Sex</td>
<td>Percent Female (n)</td>
<td>88% (14)</td>
<td>77% (10)</td>
<td>100% (11)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Percent Male (n)</td>
<td>12% (2)</td>
<td>23% (3)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SSc duration (months)</td>
<td>Median (range)</td>
<td>84 (7-348)</td>
<td>24 (0-96)</td>
<td>96 (1-216)</td>
<td>0.0349</td>
<td>&gt; 0.9999</td>
</tr>
<tr>
<td>ILD duration (months)</td>
<td>Median (range)</td>
<td>-</td>
<td>12 (0-43)</td>
<td>60 (5-108)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FVC (% predicted)</td>
<td>Median (range)</td>
<td>106.9 (91-154)</td>
<td>71 (49-115)</td>
<td>80 (53-93)</td>
<td>0.0001</td>
<td>0.0012</td>
</tr>
<tr>
<td>Medications (%)</td>
<td>Immunosuppressors</td>
<td>2 (13)</td>
<td>6 (46)</td>
<td>4 (36)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Corticosteroids</td>
<td>2 (13)</td>
<td>3 (19)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1. Study subjects characteristics

Identification of differentially expressed proteins between progressive ILD and stable ILD

Based on the 1129 proteins from the SOMAscan assay data from 16 lSSc subjects, 11 stable ILD and 13 progressive ILD; we identified 50 proteins that were differentially regulated between progressive ILD and stable ILD (Wilcoxon rank sum test False Discovery Rate (FDR) <0.1). Out of the 50 proteins identified, 25 proteins were upregulated (significantly increased) and 25 proteins were downregulated (significantly decreased) in the progressive ILD group compared to stable ILD group. The
discriminatory power of these differentially expressed proteins was tested using unsupervised hierarchical clustering. As shown in Figure 2, the spectral counts of these proteins resulted in near complete separation of the progressive cases from the stable cases with only one exception where one stable case was clustered with the progressive cases. There was visible overlap between the IISc cases and the stable cases. Four upregulated proteins (Activin-A, PARC, Calpain-1, and TARC) and three downregulated (Kallistatin, Adiponectin, and Coagulation Factor V) were selected based on clustering analysis and statistical significance.
**Figure 2 Heatmap.** Unsupervised hierarchical clustering was carried out on the basis of the expression pattern. The differentially expressed proteins were linked together according to their expression (dendogram on the left) LSSc, stable ILD and progressive ILD were also clustered (dendogram on top). Protein expression intensities were standardized between -3.0 (blue) and 3.0 (red).
Serum levels of biomarkers

Median values for serum Coagulation Factor V were 10311 RFU (6396 to 12260) for lSSc, 7161 RFU (2140 to 8296) for progressive ILD patients and 9646 RFU (6510 to 11941) for stable ILD patients. Progressive ILD patients demonstrated significant lower Coagulation Factor V levels than lSSc patients (p=0.001) and stable ILD patients (p=0.0016) (Figure 3A).

![Coagulation Factor V](image)

**Figure 3A.** Comparison of Serum Coagulation Factor V in lSSc, stable ILD (STA) and progressive ILD (PRO).

Median values for serum Calpain-1 were 12759 RFU (9904 to 15498) for LSSc, 15129 RFU (11091 to 24561) for progressive ILD and 11876 RFU (10271 to 14249) for stable ILD patients. Progressive ILD patients had higher levels of Calpain-1 than LSSc patients (p=0.0015) and stable ILD patients (p=0.0005) (Figure 3B).
Figure 3B Comparison of Serum Calpain-1 in LSSc, stable ILD (STA) and progressive ILD (PRO).

Median values for serum PARC in LSSc patients were 4979 RFU (2356 to 14943), 9140 RFU (4319 to 14476) in progressive ILD patients and 5286 RFU (3345-8628) for stable ILD patients. Progressive ILD patients had higher levels of PARC than LSSc patients (p=0.0056). Serum levels of PARC was not found to be significant between Progressive and stable ILD patients (p=0.0938) (Figure 3C).
Figure 3C Comparison of Serum PARC in lSSc, stable ILD (STA) and progressive ILD (PRO).

Results of the comparisons of other evaluated proteins are shown in Table 2

<table>
<thead>
<tr>
<th>Analyte</th>
<th>LSSc</th>
<th>Progressive</th>
<th>Stable</th>
<th>LSSc vs PRO</th>
<th>LSSc vs STA</th>
<th>PRO vs STA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coagulation Factor V</td>
<td>10311 (6396-12260)</td>
<td>7161 (2140-8296)</td>
<td>9646 (6510-11941)</td>
<td>0.001</td>
<td>&gt; 0.9999</td>
<td>0.0016</td>
</tr>
<tr>
<td>Adiponectin</td>
<td>4127 (1276-7859)</td>
<td>1698 (1069-4657)</td>
<td>2489 (1679-8877)</td>
<td>0.0015</td>
<td>0.6934</td>
<td>0.1294</td>
</tr>
<tr>
<td>TARC</td>
<td>1329 (823-2828)</td>
<td>2696 (598-6800)</td>
<td>1276 (621-2737)</td>
<td>0.036</td>
<td>&gt; 0.9999</td>
<td>0.0393</td>
</tr>
<tr>
<td>PARC</td>
<td>4979 (2556-14943)</td>
<td>9140 (4319-14476)</td>
<td>5286 (3345-8628)</td>
<td>0.0056</td>
<td>&gt; 0.9999</td>
<td>0.0938</td>
</tr>
<tr>
<td>Calpain-1</td>
<td>12759 (9904-15498)</td>
<td>15129 (11091-24561)</td>
<td>11876 (10271-14249)</td>
<td>0.0015</td>
<td>&gt; 0.9999</td>
<td>0.0005</td>
</tr>
<tr>
<td>Kallistatin</td>
<td>50529 (31165-64497)</td>
<td>40940 (25372-44877)</td>
<td>50827 (38456-58878)</td>
<td>0.001</td>
<td>&gt; 0.9999</td>
<td>0.004</td>
</tr>
<tr>
<td>Activin-A</td>
<td>7049 (4010-10893)</td>
<td>11156 (6916-16491)</td>
<td>6961 (4027-11761)</td>
<td>0.0011</td>
<td>&gt; 0.9999</td>
<td>0.0269</td>
</tr>
</tbody>
</table>

Table 2. Comparison of serum biomarkers as measured by Krustal-Wallis Test. Data are presented as median (range)

Evaluation of the selected biomarkers

Each of the seven proteins were examined for their individual ability to discriminate between subjects with progressive ILD and those with stable ILD. In analytes where subjects with progressive ILD have higher protein levels than those with stable ILD, the
area under the curve (AUC) represents the probability that a randomly selected progressive ILD patient will have a higher test result than randomly selected stable ILD patients. 95% Confidence intervals for AUC were calculated using n=1000 bootstrap replicates. Receiver-operating characteristic (ROC) curves were developed to evaluate and compare the prognostic power of our seven selected biomarkers. Positive predictive value (PPV) and negative predictive values (NPV) were also calculated. PPV measures the proportion of positive test that are true positives and represent the presence of progressive ILD whereas the NPV measures the proportion of negative test that are true negatives and represent the absence of progressive ILD.

When progressive ILD patients were compared to patients with stable ILD, a Coagulation Factor V cut-off level of 8,424.64 RFU identified the presence of progressive ILD with a sensitivity of 99%, specificity of 82%, PPV of 87%, NPV of 99% and area under the curve (AUC) of 0.93 (95% CI 0.82-1).

A Calpain I threshold of 13,717.55 RFU identified the presence of progressive ILD with a sensitivity of 92%, specificity of 91%, PPV of 92%, NPV of 90% and AUC of 0.93 (95% CI 0.82-1).

Using a PARC threshold of 8,884.0 RFU, sensitivity, specificity, and AUC for progressive ILD were 54%, 99% and 0.80 (95% CI 0.62-0.98) respectively.

Twelve out of the thirteen patients with progressive ILD, had Calpain I serum RFU measurements above the threshold. All of the progressive ILD patients had Coagulation Factor V measurements under the threshold. Therefore all of the progressive ILD patients were within the cut-off levels of Calpain I and Coagulation Factor V. Two stable ILD
patients had Coagulation Factor V measurements under the threshold, however these patients were also under the Calpain I threshold.

ROC measurements of the other evaluated proteins are shown in Table 3.

Graphical depictions of the ROC curves are shown in Figure 4. A logistic regression model was built to observe if the combination of two biomarkers increase the power of the prognostic test (Figure 5). The two proteins chosen to build the model were Coagulation Factor V and Calpain I they were chosen for combination because out of the seven selected proteins they have the highest AUC with an identical 0.93. Combining both measurements resulted in an AUC of 0.97 (95% CI 0.91-1), sensitivity of 99%, specificity of 91%, PPV of 93% and NPV of 99% that was superior to the individual measurements of the proteins (Table 4).

<table>
<thead>
<tr>
<th>Analyte</th>
<th>AUC</th>
<th>95% CI</th>
<th>Threshold</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activin-A</td>
<td>0.82</td>
<td>0.65-0.99</td>
<td>7709.95</td>
<td>0.92</td>
<td>0.64</td>
<td>0.75</td>
<td>0.88</td>
</tr>
<tr>
<td>Calpain-1</td>
<td>0.93</td>
<td>0.82-1</td>
<td>13717.55</td>
<td>0.92</td>
<td>0.91</td>
<td>0.92</td>
<td>0.90</td>
</tr>
<tr>
<td>Coagulation Factor V</td>
<td>0.93</td>
<td>0.82-1</td>
<td>8424.65</td>
<td>1.00</td>
<td>0.82</td>
<td>0.87</td>
<td>1.00</td>
</tr>
<tr>
<td>TARC</td>
<td>0.79</td>
<td>0.59-0.98</td>
<td>2040.50</td>
<td>0.77</td>
<td>0.82</td>
<td>0.83</td>
<td>0.75</td>
</tr>
<tr>
<td>Kallistatin</td>
<td>0.89</td>
<td>0.73-1</td>
<td>45863.10</td>
<td>1.00</td>
<td>0.82</td>
<td>0.87</td>
<td>1.00</td>
</tr>
<tr>
<td>PARC</td>
<td>0.80</td>
<td>0.62-0.98</td>
<td>8884.00</td>
<td>0.54</td>
<td>1.00</td>
<td>1.00</td>
<td>0.65</td>
</tr>
<tr>
<td>Adiponectin</td>
<td>0.78</td>
<td>0.58-0.97</td>
<td>1897.85</td>
<td>0.62</td>
<td>0.91</td>
<td>0.89</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Table 3. Logistic Regression Model ROC measurements
Figure 4. Analytes ROC curves

Table 4. Combined Logistic Regression Model ROC measurements

<table>
<thead>
<tr>
<th>Analyte combination</th>
<th>AUC</th>
<th>95 % CI</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calpain-1 + Coagulation Factor V</td>
<td>0.972</td>
<td>0.9123-0.99</td>
<td>0.99</td>
<td>0.909</td>
<td>0.93</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Figure 5. Combined Analytes ROC curve
DISCUSSION

In the clinical setting a prognostic biomarker is one that can separate a diseased population into groups of similar prognosis. In SSc, there is an important need for identifying prognostic biomarkers for ILD in SSc. The alarming fact that almost 90% of subjects with SSc will develop some sort of pulmonary complication including ILD highlights this need. Utilizing prognostic biomarkers in the serum of patients provides a non-invasive approach to predict the likelihood of ILD progression and allow early treatment to those at risk while avoiding unnecessary treatment for those whose disease will remain stable without treatment. Prognostic biomarkers will also contribute to the ability to enrich the population enrolled into clinical trials where at the present it may consist of a large number of subjects with stable disease. The enrollment of a large amount of subjects with stable ILD can potentially obscure the effects of potentially effective therapeutic treatments.

In this pilot study the levels of several proteins were found to be different in the serum of subjects with progressive ILD compared to stable ILD. This is the first study to use Somascan’s high throughput aptamer-based technology to evaluate the proteomic profile of systemic sclerosis patients with ILD. This type of technology provides an unbiased approach by providing efficient highly multiplexed measurements of thousands of proteins from small samples volume.

The literature surrounding the mechanistic aspects of several of these proteins make them particularly interesting in their potential utility to be prognostic biomarkers. Calcium-dependent intracellular cysteine protease (Calpain-1) shown here to be
positively associated with progressive ILD (p=0.0005), has been previously associated to play a role in other autoimmune diseases such as Rheumatoid Arthritis (RA). Furthermore, a previous study investigated whether Calpetin, a calpain inhibitor, prevented pulmonary fibrosis on bleomycin-induced mice. The group found that Calpetin prevented bleomycin-induced pulmonary fibrosis and decreased mRNA levels of genes associated with pulmonary fibrosis like IL-6 and TGF-B1 in the treated mouse lung tissues. These results seem to support the role of Calpain-1 in the progression of pulmonary fibrosis and could also suggest that it might be a good therapeutic target.

Another protein identified as downregulated in subjects with progressive ILD (p=0.0016) is Coagulation Factor V, which functions as a cofactor in the coagulation system. Recent evidence has demonstrated evidence that procoagulant signaling contributes to inflammation and fibrosis. In addition, the coagulation system interacts with the complement system where dysregulation or inhibitory functions in one or both systems are associated with clinical manifestations of systemic lupus erythematosus. The combination of Factor V and X results in the production of high quantities of thrombin a known mediator of pulmonary fibrosis in SSc-ILD and idiopathic pulmonary fibrosis (IPF). These results could suggest and support the belief that the coagulation system plays an important role in SSc-ILD and to our knowledge this is the first study to find such evidence in SSc patients with progressive ILD.

Calpain-1 and Coagulation Factor V were combined utilizing a generalized linear model to increase their prognostic power. The combined performance of these biomarkers was assessed by the area under the receiver-operator characteristics curve,
which is the most utilized measure to test the discriminatory ability of single biomarkers for binary disease outcomes. The combination of the markers produced an AUC of 0.97 (95% CI, 0.91-0.99), and a sensitivity of 99% and a specificity of 91%. Since the sample size in this study was small and the stable ILD group only included females, confounders and covariates were not included in the combined logistic regression model.

In addition to Calpain-1 and Coagulation Factor V we were able to identify other proteins that were differentially regulated between progressive ILD and stable ILD including Thymus and activation regulated chemokine (TARC) and Pulmonary and activation-regulated chemokine (PARC). TARC also known, as CCL17 is a chemokine that is usually expressed in the thymus. Chemokines are key factors that regulate the recruitment of specific immune cells into inflamed tissue. Previously TARC serum levels have been shown to be significantly elevated in patients of rheumatic disease including SSc.\textsuperscript{40} It has also been demonstrated that TARC might play a key role in the development of pulmonary fibrosis by recruiting immune cells such as lymphocytes and macrophages.\textsuperscript{41} In line with previous literature this study found that the serum levels of TARC in patients with progressive ILD were significantly elevated (p= 0.03) when compared to subjects with stable ILD. TARC also showed a good specificity value of 82% suggesting its ability to correctly identifying subjects that do not have progressive ILD. These findings support previous suggestions that TARC may be a good marker for progressive ILD.\textsuperscript{40,41}

PARC also known as CCL18 is a chemokine that has also been previously described as a potential indicator of ILD activity in idiopathic ILD and SSc-ILD.\textsuperscript{22,28–30}
Studies have shown higher levels of PARC in the lung tissue and serum when compared to controls. In our dataset we were not able to find statistical differences between the progressive ILD and stable ILD (p=0.09), however it seems to be trending towards statistically significance and a larger sample size might find a more robust difference. The small sample size of this study is a limitation, but it provides a starting point in the novel identification of potential prognostic biomarkers for ILD and new interest in combinations of analytes that have not been studied before. Another limitation is the lack of a validation cohort. The previous observation with TARC allows the current dataset to validate this analyte, but for other analytes we hope to identify another patient population as a validation cohort for the future. This cohort only included lSSc subjects who suffered from ILD, which can also be viewed as a limitation since patients with the diffuse disease can also suffer from ILD. This study is part of a larger biomarker study that assessed biomarkers for other conditions of the disease like skin progression in dSSc patients. Thus, in order to avoid any possible confounders between skin progression and ILD only lSSc subjects were selected in this study. The ideal validation cohort for the possible prognostic biomarkers identified in this study should also include subjects with dSSc in order to evaluate whether this findings also apply to this population. This would allow us to increase the power of the study and allow us to build better models to determine whether other possible confounders such as the use of immunosupresors affect the results.

For future studies instead of a binary classification of progression, the correlation of these markers with FVC as a continuous variable would be interesting. In order to
determine if particular levels of the proteins are associated with degree of decrease in pulmonary function. In addition, it would be interesting to include other PFT measurements such as the diffusing capacity of the lungs for carbon monoxide (DLCO) in order to evaluate whether the identified proteins have an effect in this measurement. Other studies could evaluate if the extent of fibrosis assessed by HRCT correlate with the proteins, which was not possible in this study due to the lack of this data.

This study is the first of its kind to use a high throughput technology to investigate the proteome of SSc and show large numbers of proteins differentially regulated between those subjects with and without progressive ILD. In particular combining the levels in the concentration of Calpain-1 and Coagulation Factor V in the serum of subjects with progressive ILD allowed for the best prediction of clinical deterioration. These findings could also help understand more of the not completely understood pathogenesis of the disease and provides other possible targets for new therapeutic options. Validation of these markers utilizing commonly used diagnostic tools such as enzyme-linked immunosorbent assay (ELISA) will be important to determine if these analytes can be evaluated using more cost-effective technology that will be practical for use in the clinical setting. In order to validate these potential prognostic biomarkers in the clinical setting, studies with a larger number of subjects will be needed to confirm their reproducibility. However, we provide early evidence of the prognostic ability of these markers to identify those subjects with progressive ILD from those with stable ILD. The validation of these analytes could potentially impact the way clinicians decide the course of treatment for patients with progressive ILD by allowing them to
provide early treatment in order to minimize or prevent further disease progression. In addition, clinical trials could benefit from the validation of these markers. These prognostic markers would allow the recruitment of those subjects who suffer from progressive disease thus enriching trials with the target population and minimizing the risk of including subjects with stable disease. The validation of these novel proteins can also help fathom the not fully understood pathogenesis of the disease and provide alternative therapeutic targets for drug or biologic development.

In this study we provided early evidence of the prognostic ability of Calpain-1 and Coagulation Factor V to identify those patients with progressive ILD from those with stable ILD.
## LIST OF JOURNAL ABBREVIATIONS

<table>
<thead>
<tr>
<th>Journal Abbreviation</th>
<th>Full Journal Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthritis Rheum</td>
<td>Arthritis and Rheumatism</td>
</tr>
<tr>
<td>BMJ</td>
<td>BMJ: British Medical Journal</td>
</tr>
<tr>
<td>Clin Exp Immunol</td>
<td>Clinical and Experimental Immunology</td>
</tr>
<tr>
<td>Clin Rheumatol</td>
<td>Clinical Rheumatology</td>
</tr>
<tr>
<td>Curr Rheumatol Rev</td>
<td>Current Rheumatology Review</td>
</tr>
<tr>
<td>Curr Opin Cell Biol</td>
<td>Current Opinion in Cell Biology</td>
</tr>
<tr>
<td>J Dermatol</td>
<td>Journal of Dermatology</td>
</tr>
<tr>
<td>J Immuno</td>
<td>Journal of Immunology</td>
</tr>
<tr>
<td>J Rheumatol</td>
<td>Journal of Rheumatology</td>
</tr>
<tr>
<td>Mol Cell Biol</td>
<td>Molecular and Cellular Biology</td>
</tr>
<tr>
<td>NEJM</td>
<td>New England Journal of Medicine</td>
</tr>
</tbody>
</table>
REFERENCES


40. Fujii H, Shimada Y, Hasegawa M, Takehara K, Sato S. Serum levels of a Th1 chemoattractant IP-10 and Th2 chemoattractants, TARC and MDC, are elevated in

CURRICULUM VITAE

JULIO C. MANTERO

Date of Birth: 1991
Place of Birth: San Juan, Puerto Rico

Professional Address

Arthritis Center
Boston University School of Medicine
72 East Concord Street, Evans 501
Boston, MA 02118
Ph. 617-638-4436
Email: jmantero@bu.edu

Education

2013-B.S. Penn State University (Biology)
2013-Present Boston University School of Medicine (Master’s Degree)

Present Occupation

2015-Present Laboratory Technician
Arthritis Center, Boston University School of Medicine

Previous positions

2014-2015 Clinical Research Assistant
Boston University Scleroderma Center of Research Translation
2012-2013 Undergraduate Research Assistant
Penn State University