A Novel Computational Approach for Scalable Biomarker Discovery in Autoimmune Diseases

Richard Chen, Grace Gao, San-He Wu, Yuanzhong Ou, Peter Schulam, Livia Casciola-Rosen, Fred Wigley, Laura Hummers, Suchi Saria

Abstract

Systemic autoimmune diseases such as scleroderma are difficult to treat because of the degree of heterogeneity in symptoms across individuals. Although skin fibrosis is the most recognized pathological feature of scleroderma, there does not exist a classification scheme that can stratify the enormous patient-to-patient variability in autoantibody profiles, disease progressions, and responses to treatments. In this study, we use a process called whole-autoantibody mapping to characterize subpopulations with similar phenotypic profiles.

Introduction

Systemic sclerosis (scleroderma) is a complex autoimmune disease affecting the circulatory system and pulmonary system, in which autoantibodies produced by B cells attack not only pathogens, but also healthy proteins, cells, and connective tissue. Scleroderma can also progress with visible skin manifestations that range from minor cases of skin involvement such limited cutaneous systemic sclerosis (lcSSC), to progressive and severe cases of skin involvement such as diffuse cutaneous systemic sclerosis (dcSSC) and fulminant systemic sclerosis (fcSSC).

Autoantibodies, which target specific cognate cellular proteins called autoantigens, are emerging as potential biomarkers that can diagnose subtypes of scleroderma. This is well-

illustrated by the three most frequently found scleroderma-specific autoantibodies: anti-topoisomerase I, anti-centromere and anti-RNA polymerase III. Anti-centromere antibodies are commonly associated with the lcSSC subtype, where as anti-topoisomerease I and anti-RNA polyermase III are commonly associated with the dcSSC subtype.

Not all cases of scleroderma, however, present with detectable skin involvement and fall into one of these subtypes. Autoantibodies against topoisomerase-1 are more prevalent in scleroderma patients with active lung disease and therefore have been used as biomarkers for identifying those at risk for greatest lung decline. Patients with antibodies against the centromere proteins frequently have disease features that include severe vascular complications, ischemic digital loss and pulmonary hypertension. In contrast, scleroderma patients with antibodies against RNA polymerase III usually have aggressive, rapidly progressive diffuse cutaneous disease and a significantly higher risk of scleroderma renal crisis. However, beyond these few exemplars, our knowledge of autoantibodies that are predictive of clinical trajectories is limited.

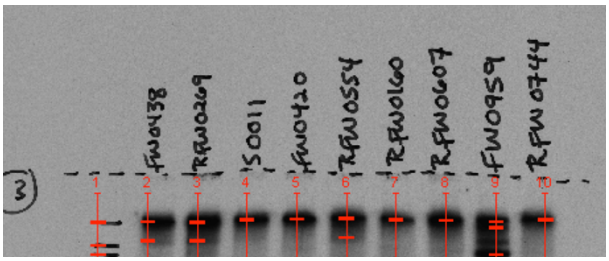
A significant bottleneck in autoantibody discovery has been the lack of a scalable process to rapidly detect candidate biomarkers. Currently, the autoantibody discovery is done using ELISA assays, which begins with a concrete hypothesis identifying one or more patient groups and a specific autoantibody that is hypothesized to be differentially expressed across these groups. This discovery process only allows for testing of one autoantibody at a time and are limited to only known molecular specificities.

We propose an alternate pipeline. We use a process called immunoprecipitation for Whole-Antibody Mapping. This process identifies the constellation of bands associated with autoantibodies present in an individual's sera. In prior work, we developed a novel clustering algorithm called the "Probabilistic Subtyping Model" (PSM) that clusters similar trajectories of

longitudinal data over time. In a group of 567 patients, we clustered a measurement for lung failure called percentage of forced vital capacity (pFVC) and identified 3 subgroups: stable, slow-declining, and fast-declining. We then identify the bands that are predictive of the different phenotypic profiles. Finally, the specific autoantibodies associated with these bands are identified (see future work) for developing new clinical assays that are of diagnostic utility.

Method

Whole-Autoantibody Mapping



In this section, we describe a procedure called Whole-Autoantibody Mapping, which maps the whole autoantibody cassette of each patient and rapidly detects the presence of any autoantibodies. At the beginning of Whole-Autoantibody Mapping, we used a process called Immunoprecipitation, in which we mixed each patient's sera with lysates. The autoantibodies present in the serum would bind to their cognate protein to form an immunoprecipiate, which can be separated by molecular weight using gel electrophoresis. The autoantibodies present in each patient's serum are read out as bands on an X-ray film, in which each lane represents data from one patient, and each band represents a specific autoantibody that was detected. The bands on the X-ray film can then be scanned by densitometry to obtain a trace of the whole-autoantibody cassette of each individual. Peaks with approximately the same trace positions and Relative Fronts (RF) values across individuals correspond to shared autoantibodies. We performed Whole-

Autoantibody Mapping to obtain the whole-autoantibody cassette of each individual across a cohort of 567 patients.

Unsupervised Learning

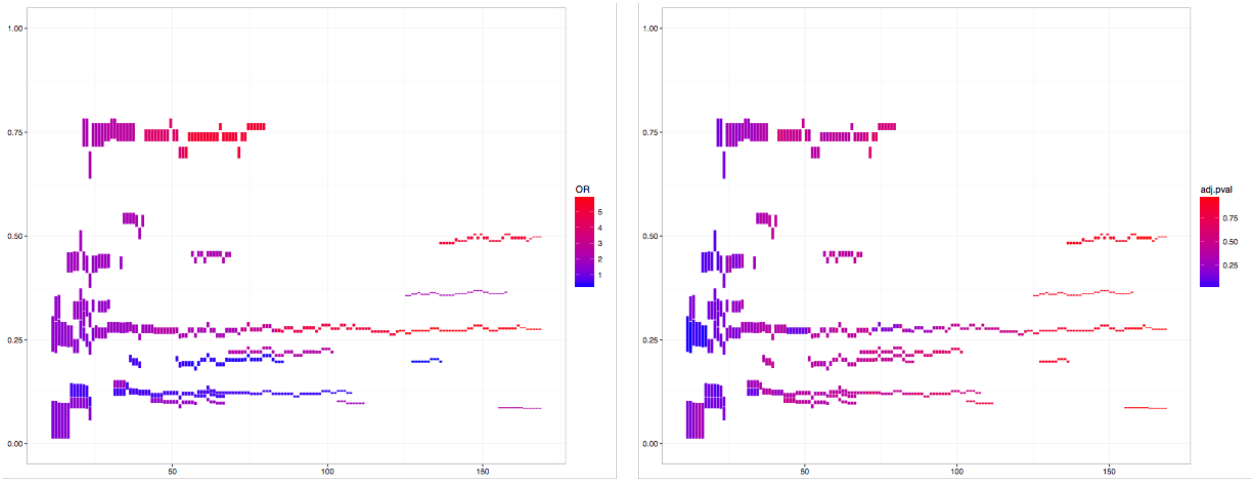
In obtaining the whole-autoantibody cassette of each patient, we realized that there was ambiguity in determining which bands were truly unique across patients. For example, in figure 1, visually, all of the patients seem to share the same first band. However, the trace positions of these bands can vary from 0.01 - 0.0001 RF, making it difficult to classify these bands as all being the same autoantibody. This variance can be explained as either noise in the densitometry, or the actual presence of distinct autoantibodies.

To determine the number of unique bands in our cohort of 567 patients, we used Jenk's Natural Break Optimization to cluster similar bands together. As little as 10 clusters and as many as 160 clusters were created to bin the data set, which each partition representing a unique band/autoantibody in that cluster. We used multiple hypothesis testing to test whether or not each partition/unique band across all clusters was predictive of the fast declining pFVC subtype found using PSM.

Results & Discussion

In this section, we conducted preliminary exploratory analysis to find candidate bands that were positively predicting of fast declining pFVC. In the adjusted P-value plot in figure 2, bands found in the 0.225-0.261 RF range were always significant across all clusters of the data. Bands around 0.75 RF were found to be only significant with 10-75 partitions, which might indicate weak

interaction effects amongst autoantibodies that are closely identical in molecular weight and shape. Bands around 0.375 RF and 0.50 RF were found to be only significant with greater than 125 partitions, which might indicate the minimum number of unique autoantibodies found in Whole-Autoantibody Mapping.



In figure 3, for patients that possessed bands in the 0.225-0.231 and 0.363-0.366 RF ranges, they were more than twice as likely to have a fast declining phenotypic profile of pFVC.