

# A Clinically and Biologically Based Subclassification of the Idiopathic Inflammatory Myopathies Using Machine Learning

Simon W. M. Eng,<sup>1</sup> Jeannette M. Olazagasti,<sup>2</sup> Anna Goldenberg,<sup>3</sup> Cynthia S. Crowson,<sup>2</sup> Chester V. Oddis,<sup>4</sup> Timothy B. Niewold,<sup>2</sup> Rae S. M. Yeung,<sup>1</sup> and Ann M. Reed<sup>5,\*</sup>

**Objective.** Published predictive models of disease outcomes in idiopathic inflammatory myopathies (IIMs) are sparse and of limited accuracy due to disease heterogeneity. Computational methods may address this heterogeneity by partitioning patients based on clinical and biological phenotype.

**Methods.** To identify new patient groups, we applied similarity network fusion (SNF) to clinical and biological data from 168 patients with myositis (64 adult polymyositis [PM], 65 adult dermatomyositis [DM], and 39 juvenile DM [JDM]) in the Rituximab in Myositis trial. We generated a sparse proof-of-concept bedside classifier using multinomial regression and identified characteristics that distinguished these groups. We conducted  $\chi^2$  tests to link new patient groups with the myositis subtypes.

**Results.** SNF identified five patient groups in the discovery cohort that subdivided the myositis subtypes. The sparse multinomial regressor to predict patient group assignments (areas under the receiver operating characteristic curve = [0.78, 0.97]; areas under the precision-recall curve = [0.55, 0.96]) found that autoantibody enrichment defined four of these groups: anti-Mi-2, anti-signal recognition peptide (SRP), anti-nuclear matrix protein 2 (NXP2), and anti-synthetase (Syn). Depletion of immunoglobulin M (IgM) defined the fifth group. Each group was associated with one subtype, with adult DM being associated with anti-Mi-2 and anti-Syn autoantibodies, JDM being associated with anti-NXP2 autoantibodies, and adult PM being associated with IgM depletion and anti-SRP autoantibodies. These associations enabled us to further resolve the current myositis subtypes.

**Conclusion.** Using unsupervised machine learning, we identified clinically and biologically homogeneous groups of patients with IIMs, forming the basis of an integrated disease classification based on both clinical and biological phenotype, thus validating other approaches and what has been previously described.

## INTRODUCTION

Idiopathic inflammatory myopathies (IIMs) encompass a heterogeneous group of chronic acquired disorders that include polymyositis (PM), adult dermatomyositis (DM), childhood myositis (predominantly juvenile DM [JDM]), myositis associated with cancer or another connective tissue disease, and inclusion body myositis (1). They are characterized by proximal muscle weakness, elevated muscle enzymes, electro-

myographic changes, and in some cases, characteristic histologic changes with cellular infiltrates on muscle biopsy. Although no standard therapeutic guidelines exist, traditional treatment has included corticosteroids and a variety of second-line immunosuppressants (2). When these treatments do not control disease, B-cell depletion with rituximab (RTX) is considered a valid therapeutic option.

Recently, promising results on the effectiveness of RTX for DM, PM, and JDM were published from the Rituximab in Myositis

This work was supported by the NIH: Mechanisms of Response and Relapse in Rituximab-treated Refractory Idiopathic Inflammatory Myopathies (AR061298-03), the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS; grants N01-AR-42273, R01-AR-061298, and R01-AR-57781), and the National Center for Advancing Translational Sciences (UL-1-TR-000137). This work was also supported by The Myositis Association (TMA), Genentech, the Cure Juvenile Myositis Foundation, and NIAMS Diversity Supplement (NIAMS R01-AR-060861). Dr. Yeung is supported by the Hak-Ming and Deborah Chiu Chair in Paediatric Translational Research.

<sup>1</sup>Simon W. M. Eng, PhD, Rae S. M. Yeung, MD, PhD: Hospital for Sick Children (SickKids), University of Toronto, Toronto, Ontario, Canada; <sup>2</sup>Jeannette M. Olazagasti, BS, Cynthia S. Crowson, PhD, Timothy B. Niewold,

MD: Mayo Clinic, Rochester, Minnesota; <sup>3</sup>Anna Goldenberg, PhD: Hospital for Sick Children (SickKids), Vector Institute, University of Toronto, Toronto, Ontario, Canada; <sup>4</sup>Chester V. Oddis, MD: University of Pittsburgh, Pittsburgh, Pennsylvania; <sup>5</sup>Ann M. Reed, MD: Duke University School of Medicine, Durham, North Carolina.

Drs. Eng and Olazagasti contributed equally to this work.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Ann M. Reed, MD, Duke University School of Medicine, Department of Pediatrics, 2301 Erwin Road, Box 3352, Durham, NC 27710. E-mail: ann.reed@duke.edu.

Submitted for publication August 15, 2019; accepted in revised form December 10, 2019.

(RIM) trial, which is a prospective randomized double-blind clinical trial (3). In this study, 200 patients with refractory myositis were randomized into two arms: one in which patients received RTX at baseline and another in which patients received RTX after eight weeks. Eighty-three percent of study patients achieved the definition of improvement, and individual core set measures improved in both RTX-treated groups. Furthermore, RTX provided a steroid-sparing effect throughout the trial. Although this trial detected no difference between the arms, its results suggest that RTX had an effect.

To maximize the effectiveness of RTX and minimize complications and cost, predictive models may identify patients with IIMs who will benefit most from RTX. Predictive models could also provide clinicians with valuable information on RTX use in IIMs. Published predictive models of disease outcomes in refractory myositis are sparse.

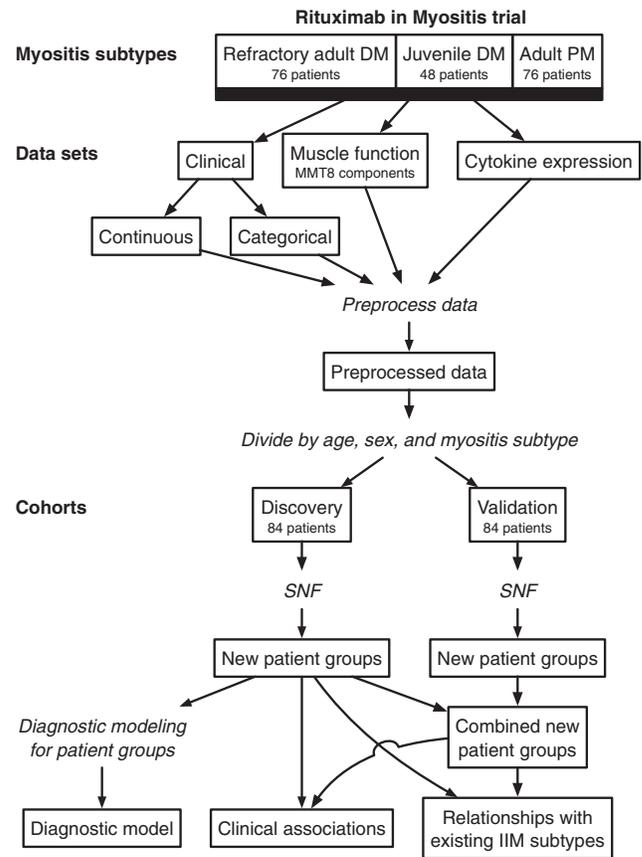
Computational methods may address the heterogeneity of IIMs and therefore aid in improving predictive power of disease course and outcomes. Previous applications of computational methods in clinical and biological contexts have demonstrated their potential for integrating diverse and heterogeneous data sets to identify patterns that distinguish patient subgroups from each other. Eng et al, in a study with newly diagnosed juvenile idiopathic arthritis, identified clinically intuitive indicators from demographic, clinical, laboratory, and cytokine expression data that produced patient groups that were more homogeneous than existing classifications (4). Wang et al developed similarity network fusion (SNF), a novel method that integrates diverse genomic and clinical data to identify homogeneous disease entities. When applied to genomic data from patients with cancer, SNF identified biologically unique patient subtypes with direct relevance to clinical outcomes (5), demonstrating the potential power of computational methods when applied in the context of IIMs.

Our refractory myositis cohort is heterogeneous, and machine learning methods may help to resolve homogeneous disease entities that predict clinical outcomes in an unsupervised approach. We sought to develop predictive models of refractory myositis with the aid of machine learning methods and to improve on our understanding of existing predictors of disease outcome using data from the RIM trial.

## METHODS

**Overview.** Figure 1 outlines the analysis conducted in this study. The analysis is detailed below.

**Study population.** The design of the RIM trial has been described in detail previously (3). The trial enrolled 200 subjects: 76 with refractory adult DM, 76 with adult PM, and 48 with JDM. Data were available for 194 of these 200 patients. The RIM trial used a randomized double-blind placebo-phase



**Figure 1.** Overview of the analysis in the study. DM, dermatomyositis; IIM, idiopathic inflammatory myopathy; MMT8, Manual Muscle Testing and a Subset of Eight Muscles; PM, polymyositis; SNF, similarity network fusion.

design of intravenous RTX in which refractory subjects were randomized to either an “early-start arm” or “late-start arm”; therefore, all subjects received RTX. Demographic, clinical features, and core set measures at baseline were similar between the early and late RTX-treated groups. There was no difference in the time to achieving the definition of improvement in any subtype (3). Baseline was defined as the time of initial treatment with RTX: week zero for the early-start arm and week eight for the late-start arm.

Informed consent and assent for participation in the RIM study including biological and data analysis.

**Clinical assessment.** Clinical assessment and disease activity measures were evaluated using a core set of measures described by the International Myositis Assessment and Clinical Studies group (6). These measures included the physician global visual analog scale (VAS), extramuscular global VAS, and Manual Muscle Testing and a Subset of Eight Muscles (MMT8). Extramuscular global VAS combines scores of a physician’s assessment of all measures, excluding muscle, and muscle disease activity VAS of the Myositis Disease Activity Assessment Tool. All study participants had their disease activity assessed at the time of study entry

and at all follow-up visits. Supplementary Table 1 lists all clinical variables used in this study.

**Autoantibody assessment.** Myositis-specific autoantibodies were measured at the University of Pittsburgh using immunoprecipitation techniques as previously described (3,7). Autoantibodies were classified into eight groups: 1) anti-aminoacyl transfer ribonucleic acid synthetases (anti-Syn), including that for histidine (anti-Jo-1); 2) anti-transcription intermediary factor 1 $\gamma$  (anti-TIF-1 $\gamma$ ); 3) anti-signal recognition particle (SRP); 4) anti-nuclear matrix protein (anti-NXP2); 5) anti-Mi-2; 6) other known myositis-related autoantibodies (anti-polymyositis-scleroderma, anti-U1 ribonucleoprotein, anti-Sjögren's-syndrome-related antigen A/B, anti-Ku, anti-small ubiquitin-related modifier 1 activating enzyme [anti-SAE], anti-U1/U2, and anticentromere antibody); 7) autoantibodies that could not be definitively identified (unidentifiable group); and 8) patients with no detectable autoantibodies. Supplementary Table 1 lists all autoantibodies in this study.

#### Measurement of serum cytokines and chemokines.

Peripheral blood serum was drawn into Serum-Separator Tubes (BD Vacutainer Blood Collection Tube; Becton) at baseline, 8 weeks, and 24 weeks after the first RTX dose. Serum cytokine levels were measured by multiplexed sandwich immunoassays (Meso Scale Discovery) (Supplementary Table 1). Samples were run in duplicate and calibrated recombinant proteins were used to generate standard curves.

**Software.** Analyses were performed using JMP 11.2.1 (JMP Statistical Discovery), R statistical software version 3.4.0 (R Foundation for Statistical Computing), and Python version 3.6 (Python Software Foundation).

**Data preprocessing and discovery and validation cohorts.** To address heterogeneity in types of measurements and the ranges their values encompass, data were first preprocessed (Supplementary Text 1 and Supplementary Text 2). The remaining 168 subjects were evenly partitioned into cohorts matched by age, sex, and myositis subtype, producing discovery and validation cohorts of 84 patients each with continuous clinical data, categorical clinical data, biological data, and muscle function data.

**Similarity network fusion.** To integrate the multiple data types together, SNF (5) was applied to preprocessed data using the *SNFtool* package, version 2.2. SNF consists of a two-step process. First, for each data type, a patient similarity matrix (PSM) was calculated, measuring how similar patients were to each other by Euclidean distances. Patient similarity networks (PSNs) were then constructed from these PSMs, with nodes as patients and weighted edges as pairwise patient similarities. In the second step, these PSNs were fused together. Patients were clustered on the fused PSN using spectral clustering, resulting in new patient groups supported by all data sets. The number of new patient groups was determined using alluvial plots showing the flow of patients across increasing numbers of groups and co-clustering probabilities (Supplementary Text 3).

**Table 1.** Patient demographics<sup>a</sup>

Characteristic	Value, by Cohort			P Value
	All Patients	Discovery	Validation	
Female patients, number of patients (%)	125 (74%)	63 (75%)	62 (74%)	1.0
Caucasian race, number of patients (%)	122 (77%)	64 (76%)	58 (69%)	0.39
Disease duration, years, median <sup>b</sup>	3.2 (0.6, 1.6, 6.7, 46)	2.9 (0.7, 1.6, 7.1, 36)	3.3 (0.6, 1.5, 6.0, 46)	0.74
Myositis subtype, number of patients (%)				0.98
• Adult polymyositis	64 (38%)	32 (38%)	32 (38%)	1.0
• Adult dermatomyositis	65 (39%)	33 (39%)	32 (38%)	0.87
• Juvenile dermatomyositis	39 (23%)	19 (23%)	20 (24%)	0.86
Physician global VAS, median <sup>b</sup>	51 (4, 35, 62, 86)	50 (7, 34, 59, 84)	52 (4, 42, 63, 86)	0.080
Muscle disease activity VAS, median <sup>b</sup>	49 (0, 30, 63, 91)	45 (0, 29, 61, 84)	51 (0, 34, 63, 91)	0.11
Extramuscular disease activity VAS, median <sup>b</sup>	27 (0, 10, 46, 84)	25 (0, 10, 46, 71)	30 (0, 8.6, 46, 84)	0.79
Failed immunosuppressive agents <sup>c</sup> , median <sup>b</sup>	2 (1, 2, 3, 6)	2 (1, 2, 3, 6)	3 (1, 2, 3.5, 6)	0.43
Prednisone dosage, mg/d, mean	21	...	...	
Autoantibody profiles, number of patients (%)				
• Anti-aminoacyl transfer ribonucleic acid synthetases	26 (16%)	11 (13%)	15 (18%)	0.52
• Anti-TIF-1 $\gamma$	18 (11%)	5 (6.0%)	13 (16%)	0.080
• Anti-SRP	21 (13%)	10 (12%)	11 (13%)	0.40
• Anti-Mi-2	23 (14%)	14 (17%)	9 (11%)	0.37
• Anti-Mj	16 (9.5%)	11 (13%)	5 (6.0%)	0.19
• Other	23 (14%)	12 (14%)	11 (13%)	0.34
• Unidentified	9 (5.4%)	4 (4.8%)	5 (6.0%)	0.26
• None	32 (19%)	17 (20%)	15 (18%)	0.84

Abbreviations: SRP, signal recognition particle; TIF, transcription intermediary factor; VAS, visual analog score.

<sup>a</sup>P values for comparisons between the discovery and validation cohorts were calculated by Fisher's exact test or Wilcoxon rank sum tests as appropriate. <sup>b</sup>Values within parentheses: minimum, 25th percentile; 75th percentile, maximum. <sup>c</sup>Includes disease-modifying antirheumatic drugs and corticosteroids.

The fused PSN, whose connections between patients have attributes such as similarity, was visualized using Cytoscape 3.5.0 (8). Starting with no connections, they were added in order of descending similarity until each patient was connected with at least one other. Patients were spatially arranged using Cytoscape's Compound Spring Embedder layout algorithm using inverse squared similarities so that similar patients were closer to each other.

**Validation.** To assess whether the patient groups were generalizable, the entire SNF analysis was rerun on the validation cohort using the same parameters as the discovery cohort, including the number of new patient groups. The resulting validation patient groups were matched with discovery groups using a greedy algorithm that first matched the discovery and validation group with the minimum Euclidean distance between centroids. Remaining groups were iteratively matched in this manner until all were matched.

To increase the power of downstream analyses, the discovery and validation cohorts were combined.

**Markers of the new patient groups.** To explore a potential bedside application of the new patient groups, a classifier was constructed to identify markers that reliably assigned patients to new patient groups. Bootstrapped multinomial regression was conducted on the discovery cohort with 2000 bootstraps and  $L_1$  regularization to predict group assignments from input data. Each coefficient (patient group and variable) in the resulting regression model was the mean across these 2000 bootstraps. Performance was evaluated by applying the resulting regression model to the validation cohort.

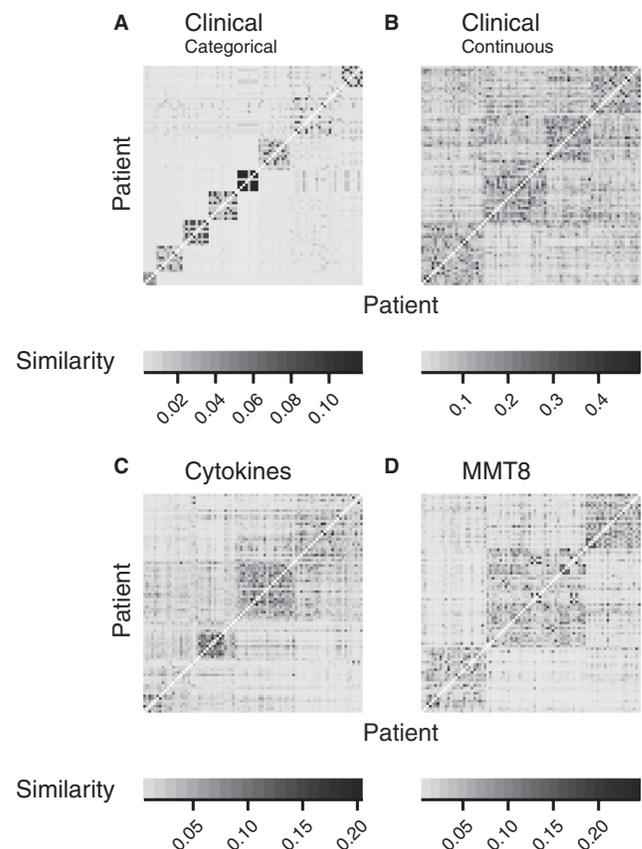
To determine whether a reduced set of variables could predict group assignments, coefficients whose 95% bootstrap confidence intervals (CIs) crossed zero were zeroed. Performance was evaluated as above.

**Distinguishing features of new patient groups.** To determine which variables distinguished new patient groups from each other, multiple linear regression was conducted to predict continuous values from groups, and multiple logistic regression was conducted to predict dichotomous variables in both the discovery and combined cohorts. Model  $P$  values were Bonferroni-adjusted to account for multiple hypothesis testing.

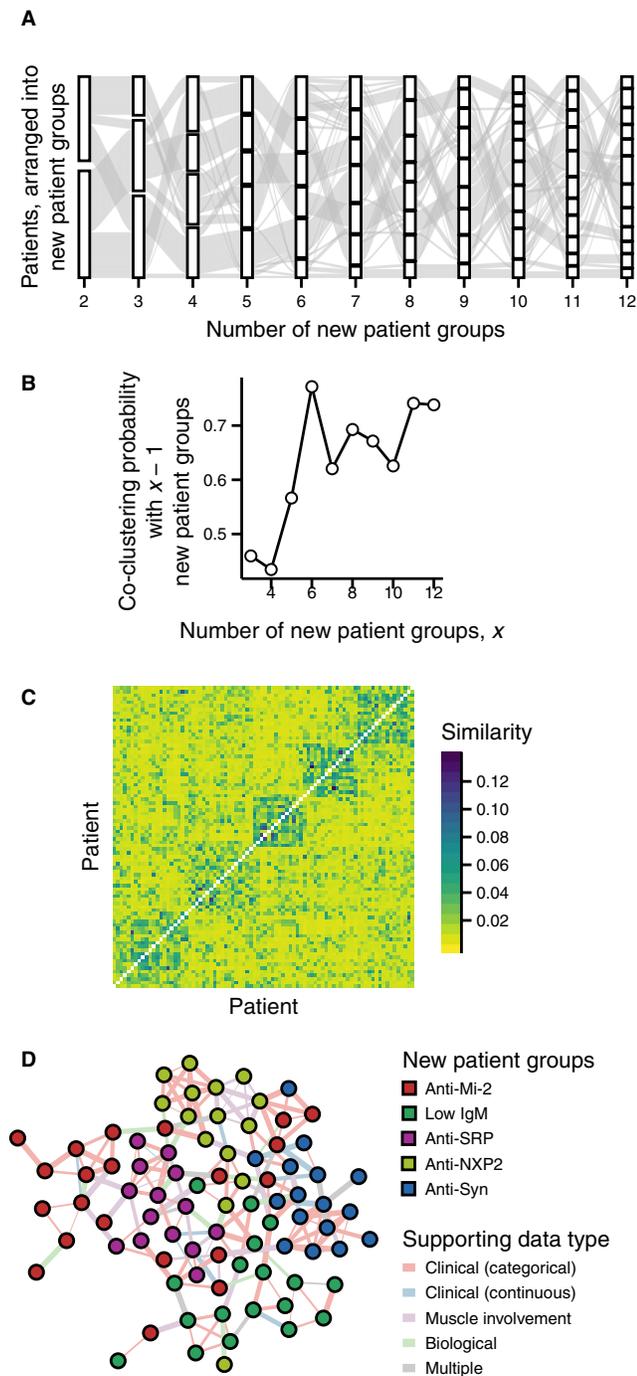
**Comparisons of new patient groups and classic myositis subtypes.** Circos 0.63 (9) was used to visualize relationships between new patient groups and myositis subtypes in both the discovery and combined cohorts. To determine over-represented combinations of new patient groups and myositis subtypes, a contingency table was first created to count patients within these combinations. A  $\chi^2$  test was then conducted on these counts (10).

## RESULTS

**Patient characteristics.** Detailed demographics, baseline disease characteristics, safety, and clinical outcomes of RIM trial participants have been previously reported (3). Cytokine data were available for 177 of 200 subjects. Among subjects included for this analysis, 122 of 168 were Caucasian (77%), and 125 of 168 were female (74%). They experienced longstanding disease (median: 3.2 years; interquartile range: 1.6-6.7 years) and highly active disease as evidenced by the Physician Global (51 mm; 35-62 mm) and Muscle Disease Activity VAS scores (49 mm; 30-63 mm) (3). Subjects failed a median of two immunosuppressive agents in addition to glucocorticoids. The mean prednisone dosage at study entry was 15.6 mg/d. Autoantibody profiles were as follows: anti-Syn in 26 (15%) patients, anti-TIF-1 $\gamma$  in 18 (11%), anti-SRP in 21 (13%), anti-Mi-2 in 23 (14%), anti-NXP2 in 16 (9.5%), other myositis-related autoantibodies in 23 (14%), unidentified autoantibodies in 9 (5.4%), and none in 32 (19%). Table 1 further describes



**Figure 2.** Distinct patient groupings within individual data sets. **A**, Heat map of similarities between pairs of patients (both axes; same order from bottom-left) for categorical clinical data. Patients were arranged by spectral clustering. Higher similarities are shaded darker (see figure legends) except on the diagonal, where patients are similar with themselves. **B**, Same as panel **A**, but for continuous clinical data. **C**, Same as panel **A**, but for biological data. **D**, Same as panel **A**, but for MMT8 components. MMT8, Manual Muscle Testing With Eight Measures.



the characteristics of included patients. The validation cohort was not different from the discovery cohort on details the discovery and validation patient cohorts (Table 1,  $P \geq 0.079$  by Wilcoxon rank sum tests and Fisher's exact tests).

**SNF identified five distinct new patient groups from clinical and biological features.** We then sought to identify new patient groups from the discovery cohort incorporating both clinical and biological data after imputing missing data with K-nearest neighbor imputation (Supplementary Text 2 and Supplementary Figure 1). PSMs for each data type (Figure 2)

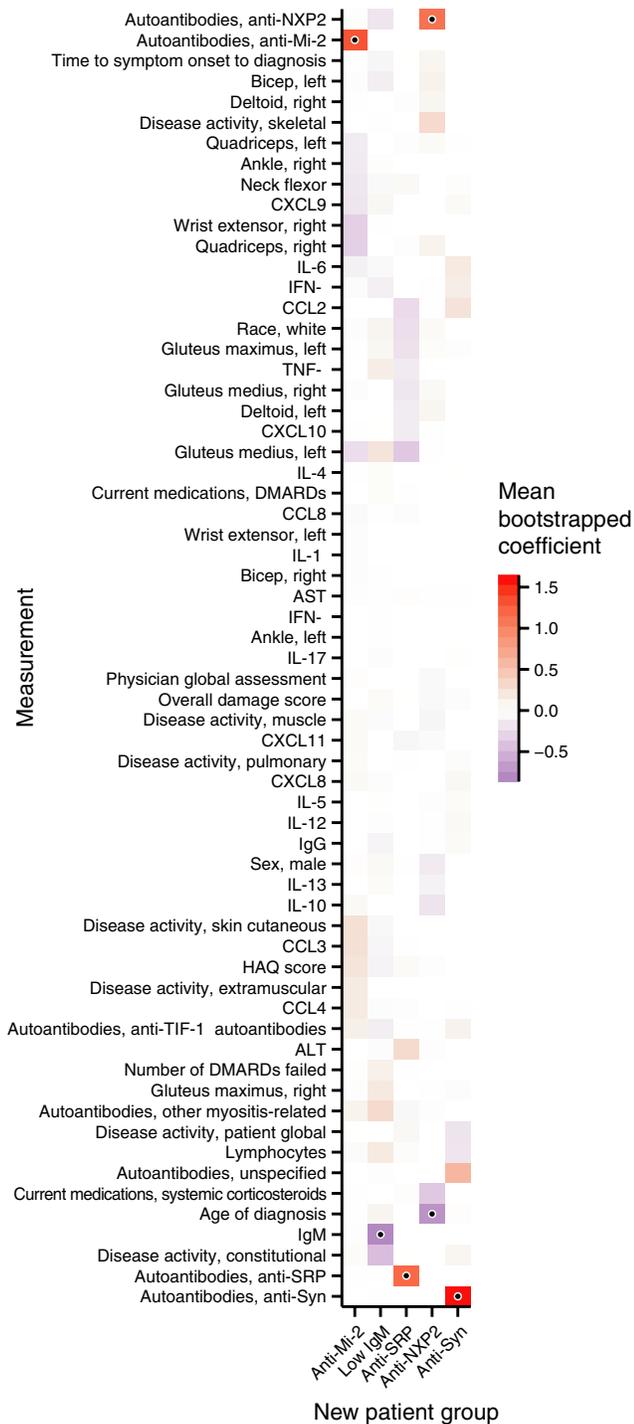
**Figure 3.** Distinct patient groupings after fusing data sets. **A**, Alluvial plot showing how patients move between new patient groups (blocks arranged vertically) as the number of groups increases (vertical groups of blocks). Gray ribbons link patients from one number of groups to another. **B**, Co-clustering probabilities (y-axis) between  $x$  and  $x - 1$  groups (x-axis). **C**, Heat map of similarities between pairs of patients (both axes; same order from bottom-left) after fusing patient similarity networks for individual data sets (Figure 1). Patients are arranged by spectral clustering. Blues indicate higher similarity (right legend) except on the diagonal, where patients are similar with themselves. **D**, Network visualization of the fused patient similarity network. Patients (nodes), colored by patient group (right upper legend), are connected by edges whose width is proportional to their similarity. Edges are colored by their strongest supporting data type(s) (right lower legend). IgM, immunoglobulin M; NXP2, nuclear matrix protein 2; SRP, signal recognition particle; Syn, synthetase.

clearly suggested at least two distinct groups per data type. Therefore, we could consider all data types when conducting SNF. Applying similar methods to the fused PSM, the five-group and six-group clusterings appeared similar to each other, having little patient movement between the five and six groups given by thick ribbons in an alluvial plot (Figure 3A). Patients clustered together most similarly with five and six groups as given by co-clustering probabilities between consecutive numbers of groups (Figure 3B). Because having six groups did not provide much more information than having five, we selected five as the number of groups. No group clearly had more or fewer patients than the others ( $\chi^2 = 1.8$ ,  $P = 0.77$  by  $\chi^2$  test; Supplementary Figure 2A). The resulting PSM and PSN supported this choice, with darker blocks comprising similar patients in the PSM (Figure 3C) and groups separating spatially in the PSN (Figure 3D). In the PSN, each data type linked at least one pair of patients with each other, as represented by colored lines. Therefore, all data types supported these groups.

Having established new patient groups in the discovery cohort, we then ran SNF independently on the validation cohort and matched the new patient groups between the two cohorts. As with the discovery cohort, no group clearly had more or fewer patients than the others ( $\chi^2 = 3.9$ ,  $P = 0.43$  by  $\chi^2$  test; Supplementary Figure 2B).

### A minimal set of autoantibodies distinguished new patient groups from each other.

To begin describing the new patient groups and to investigate a bedside application for them, we created a bootstrapped multinomial regressor to predict groups from variables used to identify them. The resulting model, trained on the discovery cohort, recovered group assignments well in the validation cohort based on receiver operating characteristic (ROC) curves and precision-recall curves (PRCs) (Supplementary Figure 3A and Supplementary Figure 3B), especially when compared with a multinomial regression model predicting myositis subtypes (Supplementary Figure 3C and Supplementary Figure 3D). Groups had areas under the ROC curve (AUROC) between 0.81 and 0.99 (Supplementary Table 2). They compared favorably to subtypes, with AUROCs of 0.34 to 0.51, suggesting



**Figure 4.** Markers for new patient groups. Heat map of mean bootstrapped coefficients (colors; right legend) for a bootstrapped multinomial regression model predicting new patient group assignment (x-axis) from measurements (y-axis). Dots (•) are coefficients whose 95% bootstrap percentile confidence interval did not cross zero. ALT, alanine aminotransferase; AST, aspartate aminotransferase; CCL, chemokine C-C motif ligand; CXCL, chemokine C-X-C ligand; DMARD, disease-modifying antirheumatic drug; HAQ, Health Assessment Questionnaire; IFN, interferon; IgG, immunoglobulin G; IgM, immunoglobulin M; IIM, idiopathic inflammatory myopathy; IL, interleukin; NXP2, nuclear matrix protein; SRP, signal recognition particle; Syn, synthetase; TIF, transcription intermediary factor; TNF, tumor necrosis factor.

group-measurement pair had a coefficient whose 95% CI did not cross zero. A small subset of measurements dominated by antibodies fulfilled this criterion (Figure 4).

As a tight set of measurements significantly predicted new patient groups, we then asked whether these antibodies alone could accurately predict these groups. The resulting pruned classifier had six antibodies with nonzero coefficients and performed favorably compared with the original unpruned classifier with AUROCs between 0.78 and 0.97 and AUPRCs between 0.55 and 0.96 (Supplementary Figure 3E, Supplementary Figure 3F, and Supplementary Table 2). Notably, the classification performance for most groups in the pruned classifier were similar to those in the unpruned classifier. Therefore, the pruned classifier could robustly predict the new groups, allowing us to name them as follows: anti-Mi-2, low IgM, anti-SRP, anti-NXP2, and anti-Syn.

**Additional clinical and biological features associated with new patient groups.**

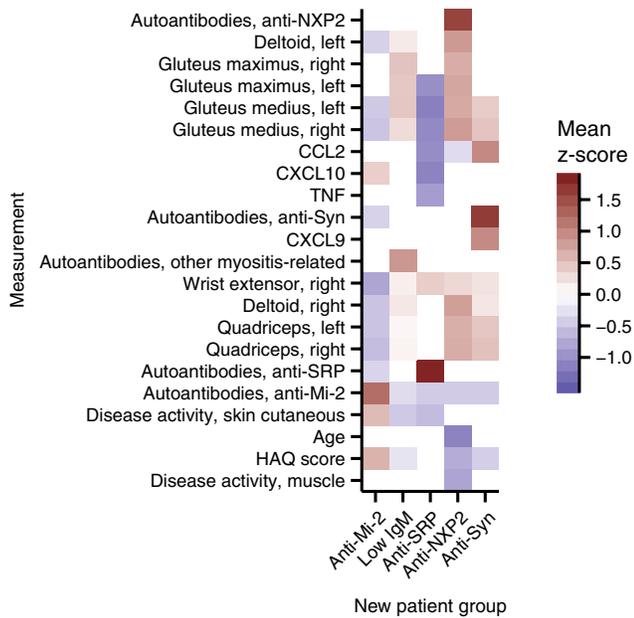
Distinct clinical and biological characteristics—beyond autoantibody profiles—were associated with the new patient groups and largely defined these groups having basal phenotypes or phenotypes appearing to extend these basal phenotypes. These phenotypes became apparent in the discovery cohort based on a heat map of distinguishing characteristics (Figure 5) and their underlying distributions (Supplementary Figure 4).

Low disease activity primarily characterized patients with anti-Mi-2 or anti-SRP autoantibodies as given by lower muscle function indicated by lower MMT8 component scores. However, anti-Mi-2 patients tended toward quadricep impairment, whereas anti-SRP patients tended toward gluteal impairment given lower mean scores for the gluteus medii. These two groups also differed in C-X-C motif chemokine (CXCL)-10 expression, with anti-Mi-2 patients having more CXCL10 and anti-SRP patients having less CXCL10.

Patients with depleted IgM had better outcomes given greater muscle function and decreased disease activity, including skin cutaneous disease activity and Health Assessment Questionnaire (HAQ) scores. In addition, these patients were enriched for other myositis-related autoantibodies and depleted for anti-Mi-2 autoantibodies, further supporting the description of the anti-Mi-2 group identified above.

that the variables as a whole predict subtypes with lower accuracy than a random classifier. Areas under PRCs (AUPRCs) for groups were between 0.64 and 0.97. AUPRCs for subtypes were between 0.24 and 0.34, further supporting the notion that the variables are unable to predict subtypes.

Having established the performance of the multinomial regressor, we then labeled each new patient group by asking which measurements were most important for classification. From the bootstrap analysis mentioned above, we determined which



**Figure 5.** Distinguishing features of new patient groups. Heat map of mean z scores (colors; bottom legend) of significantly distinguishing measurements (y-axis; Bonferroni-adjusted model  $P < 0.05$  by linear or logistic regression after Bonferroni adjustment) for each new patient group in the discovery cohort. Insignificant associations for individual groups (coefficient  $P \geq 0.05$ ) are white. Measurements were organized by agglomerative hierarchical clustering by complete linkage on Euclidean distances between mean z scores.

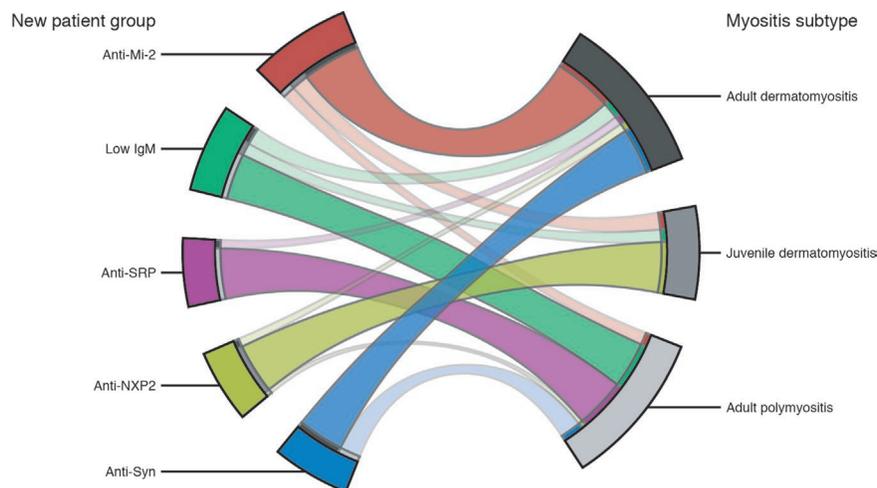
The anti-NXP2 group extended the low IgM phenotype with improved muscle function extending to the deltoids and quadriceps. These patients were also younger than all others and had the lowest disease activity measures as given by HAQ and muscle disease activity scores.

The last group, with anti-Syn autoantibodies, had decreased muscle involvement in the gluteus medii and quadriceps but had increased overall chemokine ligand 2 and CXCL9 expression.

Having previously shown that the new patient groups and their core features generalized to the validation cohort (Supplementary Figure 3), we combined the discovery and validation cohorts to provide further clarity into the defining characteristics of these groups. The combined cohort enabled us to greater resolve biological features as anti-Syn patients were associated with numerous chemokines (Supplementary Figure 5 and Supplementary Figure 6). The combined cohort allowed us to reframe anti-Mi-2 patients and anti-SRP patients in terms of the degree of muscle involvement. Both groups shared defining features in the deltoids, quadriceps, gluteus medii, and gluteus maximi, but anti-SRP patients scored lower on these measures. In the combined cohort, IgM depletion appeared to produce a more extreme phenotype as the gluteal regions no longer appeared as distinguishing features. Finally, patients with anti-NXP2 autoantibodies or who were younger continued to retain muscle function, given increased scores for the gluteal regions. The increased discriminative power, however, enabled us to identify anti-TIF-1 $\gamma$  autoantibodies as another potential marker for that group.

#### New patient groups subdivided classical myositis subtypes.

When we visualized relationships between the new patient groups and the classical myositis subtypes in the discovery cohort as a Circos figure (Figure 6), we found that all groups clearly divided the subtypes as supported by a  $\chi^2$  test ( $\chi^2 = 61$ ,  $P < 0.001$ ). Based on standardized residuals from this test (Supplementary Table 3), adult DM was associated with anti-Mi-2 and anti-Syn autoantibodies, JDM was associated with anti-NXP2



**Figure 6.** Data-driven patient groups associated with myositis subtypes. Circos figure showing relationships between new patient groups (left side; colored wedges) and myositis subtypes (right side; gray wedges) in the discovery cohort. Ribbons link patients shared between groups and subtypes. Thicker ribbons indicate more patients. Opaquer ribbons denote significantly enriched relationships, as determined by a  $\chi^2$  test, where the number of patients is higher than distributions of groups and subtypes across the entire cohort would suggest.

autoantibodies, and adult PM was associated with IgM depletion and anti-SRP autoantibodies. Patient groups rarely overlapped with other subtypes because enriched associations encompassed at least 61% of each group. These same relationships held true in the combined cohort ( $\chi^2 = 99$ ,  $P < 0.001$ ) (Supplementary Figure 7 and Supplementary Table 4), with at least 56% of patients in each new patient group participating in a significantly enriched relationship between any group and subtype.

## DISCUSSION

In this study, we stratified the IIMs into five data-driven groups using computational methods. Unique autoantibody profiles distinguished these new groups from each other despite us not explicitly defining the groups this way prior. These new groups had distinct clinical and biological features that were thematically consistent regardless of patient cohort. Furthermore, these distinguishing autoantibodies arose from a classifier that reliably assigned patients to these new patient groups, suggesting distinct underlying pathobiologies. We recognize that new subgroups have been described, ie, necrotizing myopathy, etc, and published as the 2017 EULAR/ACR criteria; however, because these subjects were enrolled prior to that time, we could not use the new criteria.

The new patient groups confirmed the diagnostic value of autoantibody profiles. The multinomial regression analysis suggested a potential bedside application for the new groups by identifying single autoantibodies that reliably classified patients into each new group, allowing us to succinctly differentiate the new groups. Age of diagnosis also defined the new groups, confirming its importance as a differentiator. Importantly, the sparse classifier retained its classification performance compared with the classifier containing all measurements. The performance of the sparse classifier has implications in bedside classification by reducing the time and cost required to make an accurate diagnosis.

Although autoantibodies almost singularly defined the new patient groups, they also served as proxies for other clinical and biological features. These new groups confirmed previously known features of the IIMs and suggested novel groupings of features tied to autoantibody profiles, contrasting benign and severe disease within the IIMs. Patients with adult DM partitioned into patients with anti-NXP2 autoantibodies, previously known as MJ, and anti-Syn autoantibodies, two new patient groups with opposing phenotypes. Patients with anti-Syn autoantibodies experienced better outcomes related to RTX treatment than those with anti-NXP2 autoantibodies because they had better muscle function in the gluteus medii and quadriceps. Meanwhile, patients with adult PM partitioned into patients with low IgM and those with anti-SRP autoantibodies. Adult PM is thought to represent a form of disease with low biological activity yet ongoing clinical disease. Patients with anti-SRP autoantibodies matched this description, having a decreased expression of several cytokines but the lowest muscle function among all new groups. This group could represent

possible muscle damage or injury with little inflammatory response such as what is seen in more necrotizing myopathy, which was not differentiated in this study. Patients with low IgM, however, had a more benign phenotype, having increased muscle function. Notably, age of diagnosis distinguished anti-NXP2 patients from all other patient groups, and JDM almost completely consisted of anti-NXP2 patients, which suggests that JDM may be distinctly different than others and has been thought to be associated with worse outcomes. NXP2 antibodies in pediatric patients have been thought to be associated with worse outcomes, including muscle contracture, atrophy, compromised functional status, and calcinosis. Most of the patients are younger, with an average age of onset around six years, are Caucasian, present with more severe symptoms of dysphonia, appear weaker, and, even though uncommon, present with gastrointestinal bleeding and ulcerations (11–14). These patients, however, had improved baseline muscle function compared with the other new patient groups, suggesting a potential subclinical signal for poor outcome that will require further study. Data were not available regarding the melanoma differentiation-associated gene 5 autoantibody, which we recognize could also confuse the results because we are working with a preexisting data set and are unable to uniformly retest all the participants.

Our analysis largely overcame a primary limitation, ie, patient numbers. From a pattern recognition viewpoint, the cohort of 168 patients would be considered small. Our use of the RIM study data was limited to patients in the trial who were treated and monitored in a standardized fashion. As a result, we could not add additional patients. Additionally, we divided the patient cohort into equally sized discovery and validation cohorts of 84 patients, which could have further reduced the power to identify and characterize patient groups. Despite these obstacles, SNF still recovered clinically and biologically meaningful large-scale features and new clinically and biologically relevant patient groups, highlighting the robustness of our analytical approach and demonstrating the applicability of SNF as a tool to study rare diseases. As such, the patient groups we have identified and characterized can be a foundation for a more detailed patient classification for the IIMs as larger, more detailed patient cohorts are enrolled.

In this study, we have described a computational approach that identified clinically and biologically homogeneous subgroups of patients with IIMs. Notably, even in a small discovery cohort of only 84 patients, SNF was able to identify homogeneous subgroups. The potential for identification of other meaningful distinguishing features using larger cohorts may enable even better predictive capabilities. Additionally, our study cohort was restricted to those patients who have refractory disease, and therefore we expected less variability among patients and thus a lower signal-to-noise ratio. Despite these challenges, SNF found signals, pointing to the excellent potential of our analytical approach in producing a biologically based classification when

applied to patients with new disease onset—where the signal-to-noise ratio would be expected to be higher. Our findings, therefore, provide foundations for an approach to precision medicine in the IIMs.

## ACKNOWLEDGMENTS

We thank all participants and investigators in the Rituximab in Myositis trial. This study would not have been possible without their contributions.

## AUTHOR CONTRIBUTIONS

Eng, Olazagasti, Yeung, and Reed were involved in drafting the article. All authors were involved in revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Reed has full access to all of the data in this study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Study conception and design.** Olazagasti, Yeung, and Reed.

**Acquisition of data.** Olazagasti, Crowson, Oddis, Niewold, and Reed.

**Analysis and interpretation of data.** Eng, Olazagasti, Goldenberg, Yeung, and Reed.

## REFERENCES

- Mammen AL. Dermatomyositis and polymyositis: Clinical presentation, autoantibodies, and pathogenesis. *Ann N Y Acad Sci* 2010;1184:134–53.
- Amato AA, Griggs RC. Treatment of idiopathic inflammatory myopathies. *Curr Opin Neurol* 2003;16:569–75.
- Oddis CV, Reed AM, Aggarwal R, Rider LG, Ascherman DP, Levesque MC, et al. Rituximab in the treatment of refractory adult and juvenile dermatomyositis and adult polymyositis: a randomized, placebo-phase trial. *Arthritis Rheum* 2013;65:314–24.
- Eng SW, Duong TT, Rosenberg AM, Morris Q, Yeung RS, REACCH OUT and BBOP Research Consortia. The biologic basis of clinical heterogeneity in juvenile idiopathic arthritis. *Arthritis Rheumatol* 2014;66:3463–75.
- Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* 2014;11:333–7.
- Isenberg DA, Allen E, Farewell V, Ehrenstein MR, Hanna MG, Lundberg IE, et al. International consensus outcome measures for patients with idiopathic inflammatory myopathies. Development and initial validation of myositis activity and damage indices in patients with adult onset disease. *Rheumatology (Oxford)* 2004;43:49–54.
- Targoff IN. Laboratory testing in the diagnosis and management of idiopathic inflammatory myopathies. *Rheum Dis Clin North Am* 2002;28:859–90, viii.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res* 2009;19:1639–45.
- Agresti A. An introduction to categorical data analysis. 2nd ed. Hoboken (NJ): John Wiley & Sons; 2018.
- Espada G, Maldonado Cocco JA, Fertig N, Oddis CV. Clinical and serologic characterization of an argentine pediatric myositis cohort: identification of a novel autoantibody (anti-MJ) to a 142-kDa protein. *J Rheumatol* 2009;36:2547–51.
- Gunawardena H, Wedderburn LR, Chinoy H, Betteridge ZE, North J, Ollier WE, et al. Autoantibodies to a 140-kd protein in juvenile dermatomyositis are associated with calcinosis. *Arthritis Rheum* 2009;60:1807–14.
- Satoh M, Tanaka S, Ceribelli A, Calise SJ, Chan EK. A comprehensive overview on myositis-specific antibodies: new and old biomarkers in idiopathic inflammatory myopathy. *Clin Rev Allergy Immunol* 2017;52:1–19.
- Tansley SL, Betteridge ZE, Shaddick G, Gunawardena H, Arnold K, Wedderburn LR, et al. Calcinosis in juvenile dermatomyositis is influenced by both anti-NXP2 autoantibody status and age at disease onset. *Rheumatology (Oxford)* 2014;53:2204–8.