



Research Paper

Adults with systemic lupus exhibit distinct molecular phenotypes in a cross-sectional study

Joel M. Guthridge^{a,b,1}, Rufe Lu^{a,b,1}, Ly Thi-Hai Tran^a, Cristina Arriens^a, Teresa Aberle^a, Stan Kamp^a, Melissa E. Munroe^a, Nicolas Dominguez^a, Timothy Gross^a, Wade DeJager^a, Susan R. Macwana^a, Rebecka L. Bourn^a, Stephen Apel^a, Aikaterini Thanou^a, Hua Chen^a, Eliza F. Chakravarty^a, Joan T. Merrill^{a,1}, Judith A. James^{a,b,1,*}

^a Arthritis and Clinical Immunology, Oklahoma Medical Research Foundation, 825 NE 13th Street, Oklahoma City, OK 73104, USA

^b Departments of Medicine and Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA

ARTICLE INFO

Article History:

Received 16 October 2019

Revised 28 January 2020

Accepted 11 February 2020

Available online xxx

Keywords:

Systemic lupus erythematosus

Biomarkers

Disease activity

Precision medicine

Disease subsetting

ABSTRACT

Background: The clinical and pathologic diversity of systemic lupus erythematosus (SLE) hinders diagnosis, management, and treatment development. This study addresses heterogeneity in SLE through comprehensive molecular phenotyping and machine learning clustering.

Methods: Adult SLE patients ($n = 198$) provided plasma, serum, and RNA. Disease activity was scored by modified SELENA-SLEDAI. Twenty-nine co-expression module scores were calculated from microarray gene-expression data. Plasma soluble mediators ($n = 23$) and autoantibodies ($n = 13$) were assessed by multiplex bead-based assays and ELISAs. Patient clusters were identified by machine learning combining K-means clustering and random forest analysis of co-expression module scores and soluble mediators.

Findings: SLEDAI scores correlated with interferon, plasma cell, and select cell cycle modules, and with circulating IFN- α , IP10, and IL-1 α levels. Co-expression modules and soluble mediators differentiated seven clusters of SLE patients with unique molecular phenotypes. Inflammation and interferon modules were elevated in Clusters 1 (moderately) and 4 (strongly), with decreased T cell modules in Cluster 4. Monocyte, neutrophil, plasmablast, B cell, and T cell modules distinguished the remaining clusters. Active clinical features were similar across clusters. Clinical SLEDAI trended highest in Clusters 3 and 4, though Cluster 3 lacked strong interferon and inflammation signatures. Renal activity was more frequent in Cluster 4, and rare in Clusters 2, 5, and 7. Serology findings were lowest in Clusters 2 and 5. Musculoskeletal and mucocutaneous activity were common in all clusters.

Interpretation: Molecular profiles distinguish SLE subsets that are not apparent from clinical information. Prospective longitudinal studies of these profiles may help improve prognostic evaluation, clinical trial design, and precision medicine approaches.

Funding: US National Institutes of Health

© 2020 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license. (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

Systemic lupus erythematosus (SLE) is characterized by remarkable clinical and pathogenic diversity, which hinders prompt diagnosis, accurate prognosis, and the optimization of therapies. Molecular profiles could be used to better understand the underlying disease processes of SLE, define features that are shared or unique in different

categorical subsets, and advance precision medicine approaches to trial designs and clinical management.

Dozens of candidate molecules with robust scientific rationale have been tested in either a broad range of SLE patients or in clinically defined subsets, such as active nephritis or cutaneous lupus patients [1]. These trials have almost universally failed to meet their pre-specified primary outcomes. Although multiple factors contribute to the failure of lupus trials [2,3], many treatments have shown efficacy in secondary endpoints or exploratory analyses [1]. Moreover, an exploratory analysis revealed differential effects of standard of care medications on the expression of genes that represent the targets for treatments currently under development. These effects varied, not only by the standard of care medication, but also by

* Corresponding author at: Arthritis and Clinical Immunology, Oklahoma Medical Research Foundation, 825 NE 13th Street, Oklahoma City, OK 73104, USA.

E-mail address: judith-james@omrf.org (J.A. James).

¹ These investigators contributed equally to this work.

Research in context

Evidence before this study

Clinical and immunologic heterogeneity in systemic lupus erythematosus (SLE) hinders effective clinical trials and treatment optimization. Recent studies have used autoantibody profiles, whole blood gene expression profiles, or immunophenotyping individually to group SLE patients into more homogenous subsets. In a large, longitudinal cohort of pediatric SLE patients, Banchereau and colleagues identified transcriptional modules that change over time with disease activity and identify clinical subsets of pediatric SLE patients. However, these approaches have not yet been tested and confirmed in adult SLE patients, nor have the intricacies of immune dysregulation been distilled into a robust tool that can be readily applied in clinical trials or in clinical care for individual patients.

Added value of this study

This study establishes an approach that identifies seven phenotypic clusters of adult SLE patients by incorporating multiple types of immunologic data through machine learning. This analysis leveraged a large, diverse, and carefully-characterized cohort to support the application of these findings to a broader population of adult patients. Disease activity and clinical manifestations varied between clusters, with certain clusters enriched for more severe disease manifestations. Clinical features that are often used to select patients for clinical trials, such as arthritis or lupus nephritis, were present across multiple clusters with distinct patterns of immune activation.

Implications of all available evidence

Multi-dimensional molecular and immunological profiles distinguish unique subsets of SLE patients that are not apparent based on clinical information, thus laying the foundation for precision medicine in lupus treatment and clinical trials.

learning approaches to cluster adult SLE patients according to their molecular phenotypes, and evaluated demographic and clinical features enriched in each cluster.

2. Methods

2.1. Patients and samples

Study procedures followed were in accordance with the ethical standards of the OMRF Institutional Review Board and with the revised Helsinki Declaration of 2000. This study was approved by the Institutional Review Board of the Oklahoma Medical Research Foundation, and all participants provided written informed consent prior to study-specific procedures. A subset of the Oklahoma Cohort for Rheumatic Disease (OCRD) at the Oklahoma Medical Research Foundation comprises 198 SLE patients meeting SLE classification by the 1997 update to the 1982 ACR criteria [21], with recruitment beginning in 2001. This subset of SLE patients from the OCRD had an average of 13 visits per subject (total cohort visits 2585, range 2–34 visits). Disease activity was measured at each visit by the modified SELENA-SLE Disease Activity Index (mSELENA-SLEDAI) [22]. SLEDAI scores ≥ 4 were considered high, and SLEDAI scores < 4 were considered low. Efforts were made to minimize bias based upon concurrent medications by the elimination of patients who had recently (within 1 year) taken rituximab, cyclophosphamide, or pulse IV steroids. In addition, samples were tested in a blinded fashion by technical personnel who were performing and initially analyzing the data. The sample size was also large enough to minimize bias. Non-autoimmune rheumatic disease controls from the Oklahoma Immune Cohort were cohort-matched for age (within 5 years), race, gender, and time of sample procurement, such that matched samples were stored at -80°C for the same length of time prior to the assays ($n = 48$) (see Supplemental Table 1 for demographics).

Blood was collected into PAXgene blood RNA tubes (PreAnalytiX, Hombrechtikon, Switzerland) for gene expression profiling. On the day of collection, the College of American Pathologists certified sample processing and biorepository core at OMRF (CAP# #9418302) isolated undiluted plasma from heparin tubes and undiluted serum from serum tubes, and froze aliquots at -80°C . Assays were performed on freshly thawed samples to maximize consistent detection of analytes, even when present in the pg/mL range.

2.2. Gene expression profiling

Total cellular RNA was isolated and purified from PAXgene tubes (PAXgene Blood RNA kit, Qiagen Inc, Valencia, CA). RNA quality and quantity were determined using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). After depletion of globin mRNA and ribosomal RNAs (Globin-Zero Gold kit, Illumina, San Diego, CA, USA), RNA was amplified, in vitro transcribed, and labeled using the Illumina Bead-Expression Kit for Illumina Human HT-12 v4.0 whole genome expression chip, and cDNA was hybridized to the chips. Chips were scanned using the Illumina iScan system. Quality control of gene expression data was performed with GenomeStudio Version 2011.1 (Illumina) per manufacturer protocol. Background-subtracted expression data were log₂ transformed and normalized with the rank-invariant method using the lumiR package [23]. System-based modular co-expression analysis was performed calculating the individual-level modular co-expression network scores (M1-M6) for each patient relative to the aggregate of controls, using second generation modular frameworks as described [18,24–26]. These modular scores were then used as individual variables in subsequent analyses described below.

underlying patient immune phenotype [4]. These observations suggest that a systematic approach is needed to address lupus heterogeneity by subdividing patients into subsets with shared immunologic and genomic characteristics.

Historical approaches to address heterogeneity have employed autoantibody profiles [5–7], clinical patterns [8–10], whole blood gene expression profiles [11–13], or immunophenotyping [14] individually to group SLE patients into more homogenous subsets; however, these approaches individually have not yet been very useful in sorting through the complex array of data acquired in clinical trials [15–17] or in developing optimized clinical care. In a large, longitudinal cohort of pediatric SLE patients, Banchereau and colleagues [18] identified transcriptional modules that change over time with disease activity and identify clinical subsets of pediatric SLE patients. In addition, our group has identified soluble mediators, including cytokines, chemokines, adhesion molecules, and soluble receptors that associate with risk of SLE disease flare [19,20]. Although the specific pathways that distinguish an impending disease flare differ somewhat between European American and African American cohorts and likely vary between individual patients, an assay that simultaneously surveys multiple immune pathways accurately identified SLE patients at higher risk of future enhanced disease activity [19,20].

Together, these findings demonstrate the value of high-dimensional data for understanding the diversity of pathogenic mechanisms in SLE and the relevance of molecular phenotypes to clinical studies and patient care. This cross-sectional study used machine

2.3. Autoantibody detection

Serum antinuclear autoantibodies were measured by indirect immunofluorescence against Hep-2 cells in the College of American Pathologists -certified Morris Reichlin Clinical Immunology Laboratory at OMRF (CAP# 2036101) [27]. ANAs present at $\geq 1:120$ dilution were considered positive. Autoantibodies were also detected using bead-based, multiplex assays on a Bio-Rad BioPlex 2200 platform (Bio-Rad Technologies, Hercules, California, USA) as previously described [28]. Antibodies against chromatin, ribosomal P, Sm, SmRNP, nRNP composite (nRNP A and/or nRNP 68), Ro/SSA composite (52 kDa Ro and/or 60 kDa Ro), La/SSB, centromere B, Scl-70, and Jo-1 were quantified as an antibody index (AI) value based on the fluorescence intensity of each of the autoantibody specificities, with a manufacturer-recommended positive cutoff of $AI \geq 1$ (range 0–>8). Anti-dsDNA was quantified in IU/mL with a manufacturer-recommended positive cutoff of 10 IU/mL.

2.4. Soluble mediator detection

Plasma levels of B lymphocyte stimulator (BlyS) and a proliferation-inducing ligand (APRIL) were measured by enzyme linked immunosorbent assays (ELISAs) per manufacturer protocol (Human BAFF/BlyS/TNFSF13B Quantikine ELISA, R&D Systems, Minneapolis, MN, USA; and Human TNFSF13/APRIL ELISA, eBioscience/Affymetrix, San Diego, CA).

Plasma levels of other soluble mediators were assessed using a custom, multiplex panel (ProcartaPlex, Thermo Fisher Scientific/Invitrogen) on the Bio-Rad Bioplex 200 Luminex xMAP plate reader (Bio-Rad Technologies) as previously described [19]. This panel uses an optimized and validated multiplex design and pre-purchase quality controls to maximize assay specificity and sensitivity, particularly for analytes present in low quantities. A bridge control serum sample was included on each plate (Collect human AB serum, Cat#2931949, Lot#Q8823, MP Biomedicals, Solon, OH, USA) to control for plate-to-plate variation of soluble mediator assays. The mean inter-assay coefficient of variance (CV) of these assays (10.5%) was within that previously shown for bead-based assays [29]. Limit of blank, limit of detection, and limit of quantification were determined and used for quality control as previously described [30]. Analytes with >60% of samples below the limit of detection were excluded from subsequent analyses. Analytes passing quality control included IFN- α , IFN- γ , IP10, MCP-1, MIG, MIP-1 α , MIP-1 β , TRAIL, TWEAK, sCD-40 ligand, TNFR I, TNFR II, IL-10, IL-17A, IL-2, IL-2RA, IL-12p70, IL-21, IL-1 α , ICAM-1, and SCF. Concentrations were interpolated from 5-parameter logistic nonlinear regression standard curves, or assigned a value of 0 if a sample was below the limit of detection.

2.5. Statistical analysis

Variables with <60% missing data were retained for univariate analyses. Quantitative variables were compared by Kruskal Wallis test, and categorical variables by Chi-square or Fisher's exact test, as appropriate, with Bonferroni adjustment for multiple corrections. No imputation was performed. Instead, samples with missing data were excluded from analyses requiring complete data, and retained for other analyses. Additionally, four outlier samples were removed from downstream clustering analyses. After these criteria, 290 samples from 194 patients were used for final analyses of molecular clusters. All variables including modular co-expression scores, autoantibody and cytokine concentrations were used for univariate, Pearson correlation and multivariate analysis in R 3.3-2.

To identify patient clusters using the informative molecular and cytokine variables, clustering and regression were performed using the unsupervised randomForest module (version 4.6–12) in R (<https://cran.r-project.org/>) with $mtry =$ the square root of the

number of variables, $ntree = 2000$, and dissimilarity matrix defined as $\sqrt{1-similarity}$, where the similarity matrix was stabilized by averaging similarities generated by 100 random forest unsupervised clustering models. The dissimilarity matrix was then reduced to three principle components using the t-Distributed Stochastic Neighbor Embedding (t-SNE) R package to minimize information loss compared to conventional dimensional reduction analysis. The first two principle components were used for K-mean clustering with $n-max$ of 20 using pamK R package.

2.6. Data sharing

Microarray data will be available through GEO (accession number GSE138458). Other data generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

3. Results

3.1. Characteristics of the study cohort

From the Oklahoma Cohort for Rheumatic Disease, this study identified 196 unique SLE patients with at least one sample that passed predefined quality control metrics for molecular phenotype data. Of these, 98 patients provided samples at two visits, which mainly had different disease activity scores (Supplemental Table 1). Of the 196 participants in this study, 89.8% were female. The racial composition was 46.9% European American, 23.0% African American, 12.2% American Indian, 11.2% Hispanic, 5.6% Asian, and the rest more than one race (Supplemental Table 1). All patients in this study were ANA positive by medical record review.

3.2. Correlations among co-expression module scores, soluble mediator levels, and SLEDAI scores

Transcriptional co-expression module scores were calculated from Illumina microarray gene expression data for previously defined immune pathway related modules [24]. Module scores showed the expected correlations (p -values <0.05) with other modules and with soluble mediator levels based on previously established immunological pathways. For example, inflammation module scores significantly correlated with each other and with the neutrophil module score. Interferon modular scores strongly correlated with one another and with levels of interferon associated soluble mediators, such as IFN α , IFN γ , IP-10 (CXCL10), MIG, BlyS, and TNFR II. Interferon module scores also significantly correlated with inflammation module scores, as well as with levels of inflammatory mediators, such as MCP-1, MIP-1 α , MIP-1 β , IL-1 α , and other cytokines (e.g. IL-2RA, IL-21, IL-12p70, IL-2 and IL-10 (Supplemental Fig. 1). B cells module scores moderately correlated with TRAIL, sCD-40 ligand, IL-21, and stem cell factor, but negatively correlated with IL-10 and TNFR I (Supplemental Fig. 1).

Total SLEDAI scores positively correlated with interferon ($r > 0.2$) and plasma cell ($r = 0.25$) module score in this adult lupus collection (Supplemental Fig. 1). Levels of IFN α , IP-10, and IL-1 α also correlated with SLEDAI scores (Supplemental Fig. 1; p -value <0.05; r range from 0.11 to 0.27). Together, these correlations demonstrate the internal validity of these datasets, the expected coordinated regulation of the expression of these markers, and correlation with clinical disease activity.

3.3. Expression modules and soluble mediators stratify SLE patients into seven subsets

To enable more precise patient stratification for future clinical studies, clinical trial designs, potential prognosis and improved clinical care, seven patient subsets with distinct molecular phenotypes were identified by k-means clustering with a t-SNE reduced random

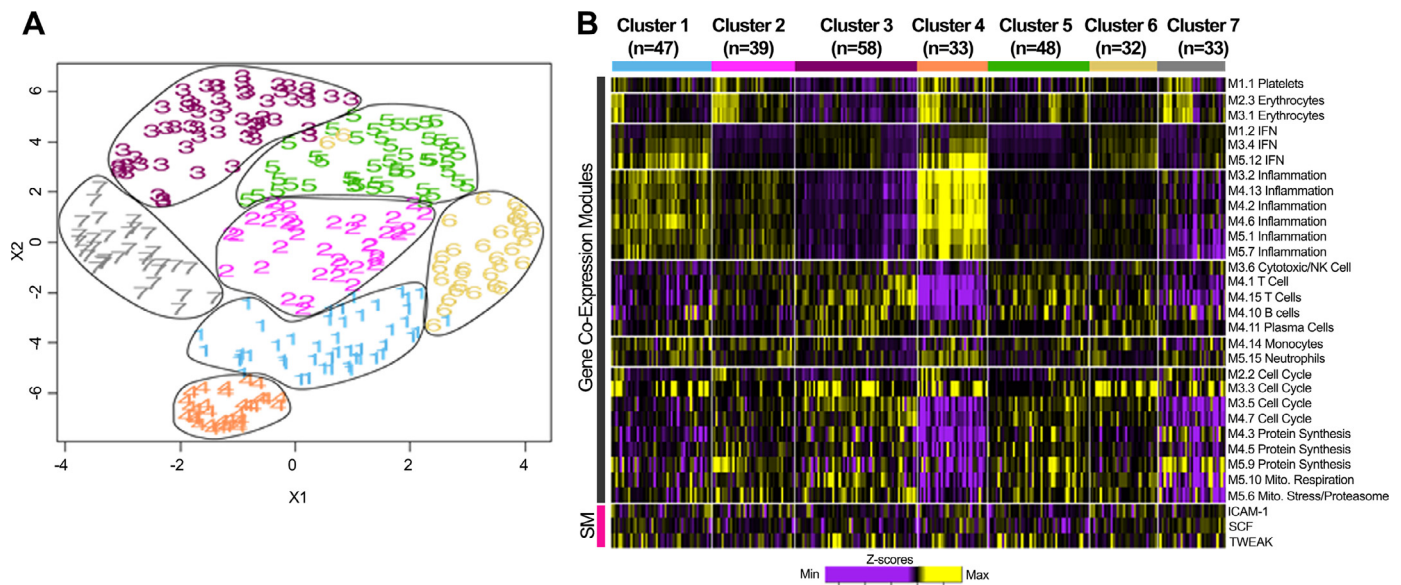


Fig. 1. Gene expression modules and soluble mediators stratify SLE Patients into seven molecular phenotypic subsets. (A) Unique patterns of gene co-expression modules and soluble mediators (SM) distinguished seven phenotypic clusters of SLE patients, indicated by colored numbers in the plot. X1 and X2 indicate the top two principal components defined by tSNE on a random forest dissimilarity matrix. These components were used in k-means clustering to identify the seven clusters. (B) The heat map presents the informative gene expression module or soluble mediator and each column is a patient. Colors indicate row z-scores, from purple (low) to yellow (high). The seven clusters are color coded as in other figures, and the number of samples in each cluster are indicated.

forest dissimilarity matrix of soluble mediators and previously defined co-expression modules [18,24–26] (Fig. 1(A)). Three clusters (Clusters 1, 4, and 6) demonstrated substantially higher IFN modular scores compared to the other subgroups. Another three clusters (Clusters 2, 3, and 5) demonstrated predominant lymphoid and monocytoïd modular scores. One cluster (Cluster 7) showed minimal activation of interferon, lymphoid, neutrophilic and monocytoïd related modules. Demographics varied between the clusters (Table 1, Supplemental Fig. 2). A higher percentage of European American patients were in Cluster 5 (24.4%) (Supplemental Fig. 2). African American patients ($n = 69$) were most often in Cluster 7 (27.5%) or Cluster 3 (24.6%), and were rarely in Cluster 1 (5.8%). American Indian patients had the highest frequency of patients in Cluster 1 (37.1%).

Clusters 1 and 4 were characterized by significant elevation of inflammation module scores, along with elevations in the interferon

module scores (Figs. 1(B) and 2(A)). Cluster 4 also had the highest neutrophil module scores (Figs. 1(B) and 2(A)) and the highest levels of soluble IP-10, MIG, APRIL, TNFR1/II, and IL-10 (Fig. 2(B)). Cluster 1 demonstrated only moderate levels of each of those variables, but had the highest levels of IL-21, IL-17A, and MIP-1 β (Fig. 2(B)). Cluster 6 had the highest levels of IL-1 α and IL-2RA. The distinguishing features based on relative expression of the gene co-expression modules and soluble mediators are shown in Fig. 2. Variations in the molecular phenotypes among these clusters suggest that different groups of lupus patients may have different active immune pathways at a given time point, or different levels of activation of particular pathways. These differences may be due to altered transcriptional regulation within cells, or differing frequencies of certain cell populations. These results suggest that distinct directed therapeutics or combination treatments may be needed to target both shared and cluster-specific immune dysregulation.

Table 1
Participant demographics and medication usage by cluster.

Cluster Number	1	2	3	4	5	6	7
Total N*	47	39	58	33	48	32	33
Female, n (% of cluster)	41 (87.2)	35 (89.7)	54 (93.1)	29 (87.9)	46 (95.8)	28 (87.5)	30 (90.9)
Age in years, mean	38.6	43.1	42.5	44.3	42.9	41.3	44.7
Race, n (% of cluster)							
European American	14 (29.8)	24 (61.5)	28 (48.3)	15 (45.5)	32 (66.7)	9 (28.1)	9 (27.3)
African American	4 (8.5)	9 (23.1)	19 (32.8)	6 (18.2)	7 (14.6)	7 (21.9)	17 (51.5)
American Indian	13 (27.7)	3 (7.7)	3 (5.2)	3 (9.1)	5 (10.4)	5 (15.6)	3 (9.1)
Asian	6 (12.8)	0 (0)	5 (8.6)	4 (12.1)	1 (2.1)	3 (9.4)	0 (0)
Hispanic	8 (17)	3 (7.7)	3 (5.2)	5 (15.2)	3 (6.3)	8 (25)	4 (12.1)
Mixed Race	2 (4.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Medication use, n (% of cluster)							
Hydroxychloroquine	30 (63.8)	31 (79.5)	40 (69)	19 (57.6)	37 (77.1)	22 (68.8)	21 (63.6)
Mycophenolate Mofetil	18 (38.3)	5 (12.8)	3 (5.2)	4 (12.1)	4 (8.3)	3 (9.4)	4 (12.1)
Azathioprine	9 (19.1)	7 (17.9)	10 (17.2)	4 (12.1)	4 (8.3)	7 (21.9)	5 (15.2)
Methotrexate	5 (10.6)	9 (23.1)	8 (13.8)	7 (21.2)	12 (25)	1 (3.1)	4 (12.1)
Cyclophosphamide	1 (2.1)	0 (0)	0 (0)	4 (12.1)	0 (0)	0 (0)	0 (0)
Rituximab	1 (2.1)	0 (0)	0 (0)	1 (3)	2 (4.2)	0 (0)	0 (0)
Steroids	27 (57.4)	10 (25.6)	22 (37.9)	27 (81.8)	16 (33.3)	14 (43.8)	15 (45.5)
Prednisone dose (mg), median (interquartile range)	0 (0, 10)	0 (0, 0)	0 (0, 0)	10 (5, 25)	0 (0, 0)	0 (0, 5)	0 (0, 10)

* Number of samples; total of 290 samples from 194 individuals.

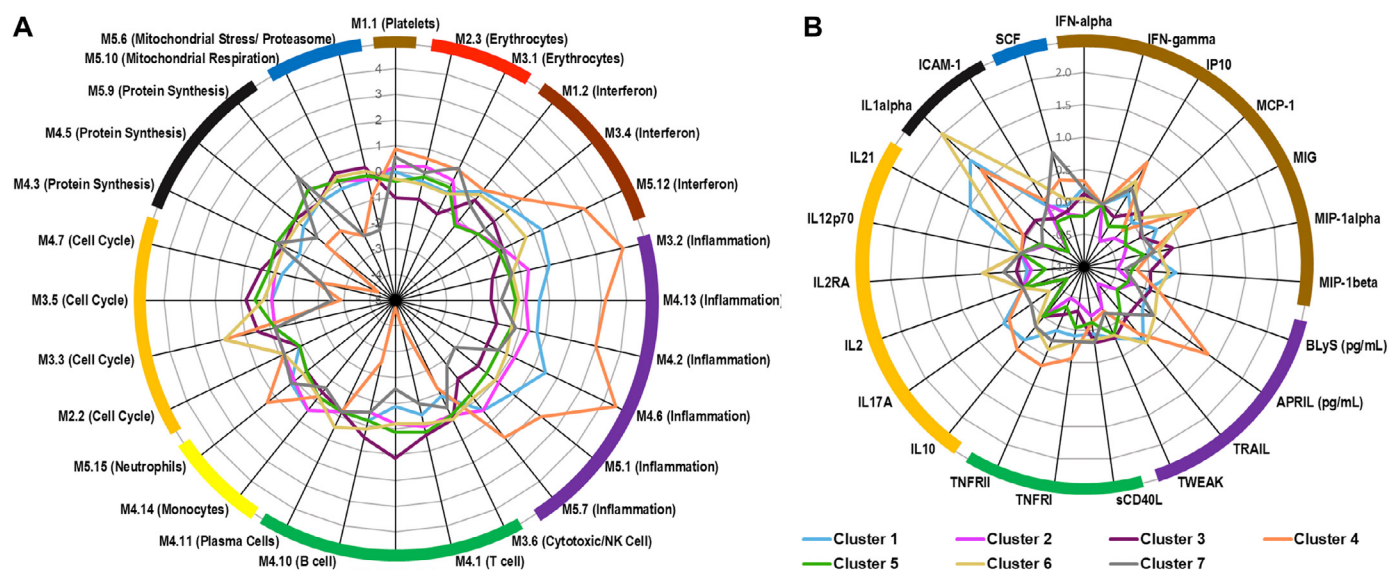


Fig. 2. Molecular profiles of seven SLE patient clusters. Radar plots show modified z-scores of relative gene expression module scores (A) and plasma soluble mediator levels (B) in each of the molecularly-defined patient clusters, indicated by colored lines as shown in the legend at bottom right. Modules and soluble mediators are grouped by function, indicated by colored arcs around each plot and labeled with the co-expression module name (A) or with the soluble mediator tested (B).

3.4. Clinical features of molecularly defined patient clusters

Every cluster included patients with higher disease activity, as well as patients with less active disease. Although there were trends towards increased SLEDAI scores in Clusters 1 and 4, there were no statistically significant differences between clusters (Fig. 3(A)). Clinical SLEDAI scores, calculated without the complement and DNA binding components, were also not statistically significant between clusters, but trended higher in Clusters 3 and 4 (Supplemental Table 2 and Supplemental Fig. 3). Furthermore, patients grouped by anti-dsDNA and complement status show substantial variability in expression modules and soluble mediators (Supplemental Fig. 4). When considering SLEDAI organ systems, each system was active in patients from multiple molecular clusters (Fig. 3(B), Supplemental Fig. 5). For example, patients with renal involvement were evenly divided between the molecularly distinct Cluster 1 ($n = 5$; 25% of renal patients), Cluster 3 ($n = 6$, 30%), and Cluster 4 ($n = 6$, 30%), with additional patients from Clusters 2 and 4 (Fig. 3(B)). Among the SLEDAI components, low complement was significantly more common in Cluster 1 than in Clusters 2 or 5 (Bonferroni-corrected $p < 0.0021$) (Fig. 3(C), Supplemental Table 3). Increased DNA binding was significantly more common in Clusters 1 and 4 than in Cluster 2 (Bonferroni-corrected $p = 0.042$ for both) (Fig. 3(C), Supplemental Table 3). Anti-dsDNA positivity and low complement levels trended higher in Clusters 1, 3, 4, and 6 (Supplemental Table 3). Other active lupus manifestations that commonly drive the selection of patients for clinical trials, such as musculoskeletal and mucocutaneous manifestations, occurred in nearly equal frequencies across all seven clusters (Fig. 3(C); Bonferroni-corrected $p > 0.05$) (Supplemental Table 3).

Cumulative ACR criteria, representing a history of organ activity, were also compared across the clusters (Supplemental Figure 6). Overall, cumulative criteria were quite similar across all of the clusters, with each criterion being represented in each cluster. Immunologic criteria were found in 75% or more of patients across all clusters, whereas renal disorder trended to more frequent in Clusters 1 and 4.

Medication use also differed somewhat between certain clusters (overall $p = 0.0001$). Frequency of steroid use was highest in Cluster 4 (Table 1). Prednisone dose was highest in Cluster 4 (all Bonferroni-corrected $p \leq 0.0092$ vs. each cluster, by Kruskal–Wallis); higher in Cluster 1 than in Clusters 2, 3, and 5 (Bonferroni-corrected $p = 0.0216$, 0.0050, and 0.0289, respectively), and higher in Cluster 7

than Cluster 3 (Bonferroni-corrected $p = 0.0428$) (Table 1, Supplemental Table 4, and Supplemental Table 5). Rates of mycophenolate mofetil use were approximately three times higher in Cluster 1 (38%) than any other cluster (5.2–13%). Hydroxychloroquine use was common across all clusters, but highest in Clusters 2 (79.5%) and 5 (77.1%) and lowest in Cluster 4 (57.6%).

3.5. Autoantibody profiles in molecularly defined SLE patients clusters

Autoantibody profiles were compared by autoantibodies against common lupus RNA- and DNA-binding proteins, such as chromatin, nRNP, Sm and Ro (Supplemental Fig. 7). Each autoantibody was present in patients across multiple clusters. For example, anti-dsDNA and anti-RNA-binding protein autoantibodies were found in clusters 1, 4, 6, and 7, which are molecularly quite different. Further, the levels of anti-dsDNA were nearly equivalent between these clusters. Clusters 2 and 5 had the lowest levels of all autoantibodies compared to the other clusters. Anti-ribosomal P autoantibodies were infrequent, but when present were most common in Cluster 4. Cluster 6 had the highest frequencies of anti-Ro and anti-La specificities (Supplemental Fig. 7).

4. Discussion

Lack of understanding of the molecular heterogeneity of SLE remains a major deterrent to optimized individualized therapy, novel target identification, therapeutic development, and clinical trial success. Indeed, dissection of disease heterogeneity has been considered one of the ten most important contemporary challenges in SLE management [31] and is highlighted as a key obstacle for development of novel treatments [2]. Three decades of failed clinical trials and limited evidence to guide the selection of standard of care treatments hinder optimal care for individual SLE patients [1]. Key roadblocks in the development of new lupus treatments include incomplete understanding of the underlying molecular mechanisms of disease, the heterogeneity of disease across poorly defined clusters of patients, high placebo response rates, the cacophony of background therapies, and other factors [2,3]. This study set forth to address the heterogeneity of lupus by applying machine learning approaches to extensive gene expression, soluble mediator, autoantibody, and clinical information in a large cohort of carefully clinically-characterized patients. These

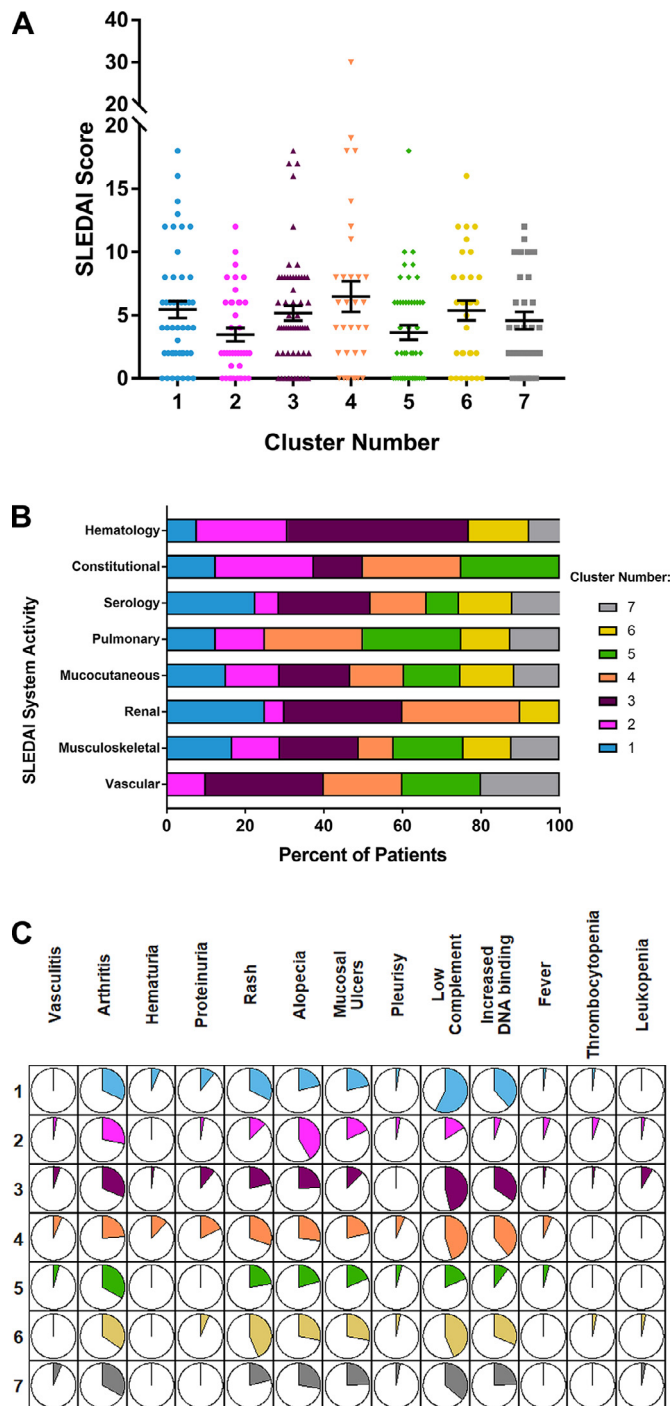


Fig. 3. Clinical phenotypes of molecularly defined SLE patient clusters, based on SLEDAI variables. (A) Mean \pm SEM SLEDAI scores in each cluster. Comparison is not significant ($p > 0.05$) by Kruskal–Wallis. (B) Each bar represents 100% of patients who have active manifestations in the given organ system, with colored segments indicating the percentage of patients from each cluster. Activity in an organ system is defined as activity in at least one of the corresponding individual components (e.g., thrombocytopenia or leukopenia in the hematologic domain; low complement or increased DNA binding in the serologic domain, etc.). (C) Frequency of SLEDAI components in each cluster. SLEDAI components not present in any patients are not shown (seizure, psychosis, organic brain, visual, cranial nerve, lupus headache, CVA, myositis, pericarditis). A pie chart showing a single line indicates 0%.

approaches identified seven phenotypic clusters of adult SLE patients that were not apparent from clinical information.

Subsetting of lupus patients has been attempted both in trials and in prognostic studies based on clinical features (such as nephritis vs.

no nephritis) or based on autoantibodies and complement. All of these approaches have taught us something, but it has long been suspected that these distinctions are somewhat incomplete, and that the next step must be to examine a more comprehensive set of immune variables. Our work and others who have utilized gene co-expression module signatures with or without the integration of soluble mediators [13,30,32,33] confirms that many features that can somewhat define patient subsets, such as autoantibodies or nephritis, are found in immunologically distinct patients who may benefit more or less from given combinations of immune modulating treatments.

Although molecular subsetting alone will not resolve every problem in the design and interpretation of lupus trials, these results may help reduce the confusion of heterogeneity in lupus based on seven characteristic patterns of gene expression and soluble mediator profiles. Important implications for treatment development are suggested, including better identification of promising treatment targets in rationally defined patient subsets, improved study of pharmacodynamic effects by sorting through their impact on molecularly similar patients, and the potential development of diagnostic tests suitable for selecting optimal treatments and monitoring individual responses based on better characterized pathologic mechanisms.

Strengths of this study include the large population of racially diverse, adult lupus patients who have donated samples when disease was both active and inactive. In addition, this project begins to integrate soluble mediators, autoantibodies, gene co-expression modules, demographic, and clinical information. The application of machine learning approaches generated seven robust clusters with strong internal validity. For example anti-dsDNA antibodies, known to be associated with nephritis, are enriched in modules that are associated with other features of nephritis.

Limitations of this study suggest major opportunities for further research. Modules evaluated were associated with a limited set of immunologic pathways. Significant additional gene expression data information is available, and novel modules within this dataset may exist which are important in adult lupus pathogenesis. The modeling being examined now might also be improved by further inclusion of genetics, epigenetics, metabolomics, lipidomics, immunophenotyping, microbiome/other environmental information, or other evolving technologies. Evaluation of larger numbers of prospectively-collected, protocolized samples across time would also be helpful to establish the stability or movement of patients between various clusters with changes in disease activity, clinical manifestations, therapeutic interventions, hormonal changes with estrus or menopause, or variation with aging. Closely related clusters, like Clusters 1 and 4, may represent two unique pathotypes, or the higher frequency of MMF use in Cluster 1 may partially suppress signatures that would otherwise be assigned to Cluster 4. Resolving this question will require studies designed to directly test the impact of therapeutics during treatment and/or disease, especially with respect to downregulation of interferon and inflammatory markers and upregulation of other pertinent immune modulators.

Synthesizing and simplifying these algorithms to a limited number of analytes may allow individual patients to be more precisely diagnosed and treated. Refinements over time will be expected, but should not delay the implementation of a good disease model which could serve clinical trial designs and future studies that will be necessary to fully integrate disparate SLE pathophysiology with prognosis for individual patients.

Declaration of Competing Interest

The authors report the following disclosures that are outside the submitted work. JMG: grants from Dxterity, Inc. CA: grants and personal fees from BMS, grants and personal fees from GSK, personal fees from AstraZeneca, grants from Exagen. MEM: grants and personal fees from Progentec Diagnostics, Inc., a patent Biomarkers for

Systemic Lupus Erythematosus Disease Activity, and Intensity and Flare (US10393739B2; EP3052193B1; assignee: OMRF) with royalties paid to Progentec Diagnostics, Inc., and a patent Biomarkers for a Systemic Lupus Erythematosus (SLE) Disease Activity Immune Index that Characterizes Disease Activity (pending patent to OMRF) licensed to Progentec Diagnostics, Inc. ET: personal fees from Neovacs, SA. JTM: other from Xencor, Incorporated, grants and personal fees from Bristol-Myers Squibb; grants and personal fees from Glaxo Smith Kline; personal fees from Lilly, Abbvie, EMD Serono, Amgen, Remegen, Servier/ILTOO, Astellas, Daiichi Sankyo, Astra Zeneca, EMD Serono, Celgene, Provention, Immupharma, Janssen, Incyte; and grants from U.S. Department of Defense. JAJ: grants and personal fees from Progentec Diagnostics, Inc.; personal fees from Abbvie, Novartis, and Janssen; grants from US Department of Defense; a patent Biomarkers for Systemic Lupus Erythematosus Disease Activity, and Intensity and Flare (US10393739B2; EP3052193B1; assignee: OMRF) with royalties paid to Progentec Biosciences, and a patent Biomarkers for a Systemic Lupus Erythematosus (SLE) Disease Activity Immune Index that Characterizes Disease Activity (pending patent to OMRF) licensed to Progentec Diagnostics, Inc.

Acknowledgments

The authors thank the clinical research teams, study participants, and regulatory personnel, as well as the OMRF Biorepository, OMRF Phenotyping Core and OMRF Clinical Genomics Center and their associated personnel. Preliminary results from this study were previously presented at the American College of Rheumatology annual meeting (Lu R, Guthridge JM, Arriens C, Aberle T, Kamp S, Munroe ME, Gross T, Dejager W, Macwana S, Roberts VC, Apel S, Chen H, Chakravarty E, Thanou K, Merrill JT, James JA. Molecular Phenotypes Associated with Clinical Disease Activity in Adult Systemic Lupus Erythematosus [abstract]. *Arthritis Rheumatol.* 2017; 69 (suppl 10)) and the Annual European Congress of Rheumatology (Guthridge JM, Lu R, Arriens C, Aberle T, Kamp S, Munroe ME, Gross T, Dejager W, Macwana SR, Bourn RL, Apel S, Chen H, Chakravarty EF, Thanou A, Merrill JT, James JA. Molecular profiles associate with clinical disease activity and inform patient subsetting in adult systemic lupus erythematosus [abstract]. *Ann Rheum Dis.* 2018; 77 (Suppl), page A1071).

Funding sources

This work was supported in part by grants from the US National Institutes of Health (U19AI082714, U01AI101934, U54GM104938, P30AR073750, UM1AI144292). The study sponsors had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; nor in the decision to submit the paper for publication.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.eclinm.2020.100291](https://doi.org/10.1016/j.eclinm.2020.100291).

References

- [1] Murphy G, Isenberg DA. New therapies for systemic lupus erythematosus – past imperfect, future tense. *Nat Rev Rheumatol* 2019;15(7):403–12.
- [2] Touma Z, Gladman DD. Current and future therapies for SLE: obstacles and recommendations for the development of novel treatments. *Lupus Sci Med* 2017;4(1):e000239.
- [3] Merrill JT, Manzi S, Aranow C, et al. Lupus community panel proposals for optimizing clinical trials: 2018. *Lupus Sci Med* 2018;5(1):e000258.
- [4] Merrill JT, Immermann F, Whitley M, et al. The biomarkers of lupus disease study: a bold approach may mitigate interference of background immunosuppressants in clinical trials. *Arthritis Rheumatol* 2017;69(6):1257–66.
- [5] Artim-Esen B, Cene E, Sahinkaya Y, et al. Cluster analysis of autoantibodies in 852 patients with systemic lupus erythematosus from a single center. *J Rheumatol* 2014;41(7):1304–10.
- [6] Kessel A, Rosner I, Halasz K, et al. Antibody clustering helps refine lupus prognosis. *Semin Arthritis Rheum* 2009;39(1):66–70.
- [7] Pacheco Y, Barahona-Correa J, Monsalve DM, et al. Cytokine and autoantibody clusters interaction in systemic lupus erythematosus. *J Transl Med* 2017;15(1):239.
- [8] To CH, Mok CC, Tang SS, Ying SK, Wong RW, Lau CS. Prognostically distinct clinical patterns of systemic lupus erythematosus identified by cluster analysis. *Lupus* 2009;18(14):1267–75.
- [9] Jacobsen S, Petersen J, Ullman S, et al. A multicentre study of 513 Danish patients with systemic lupus erythematosus. I. Disease manifestations and analyses of clinical subsets. *Clin Rheumatol* 1998;17(6):468–77.
- [10] Peschken CA, Katz SJ, Silverman E, et al. The 1000 Canadian faces of lupus: determinants of disease outcome in a large multiethnic cohort. *J Rheumatol* 2009;36(6):1200–8.
- [11] Lyons PA, McKinney EF, Rayner TF, et al. Novel expression signatures identified by transcriptional analysis of separated leucocyte subsets in systemic lupus erythematosus and vasculitis. *Ann Rheum Dis* 2010;69(6):1208–13.
- [12] Baechler EC, Batliwalla FM, Karypis G, et al. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci USA* 2003;100(5):2610–5.
- [13] Toro-Dominguez D, Martorell-Marugan J, Goldman D, Petri M, Carmona-Saez P, Alarcon-Riquelme ME. Stratification of systemic lupus erythematosus patients into three groups of disease activity progression according to longitudinal gene expression. *Arthritis Rheumatol* 2018;70(12):2025–35.
- [14] Hamilton JA, Wu Q, Yang P, et al. Cutting edge: intracellular IFN-beta and distinct type I IFN expression patterns in circulating systemic lupus erythematosus B cells. *J Immunol* 2018;201(8):2203–8.
- [15] Rovin BH, Furie R, Latinis K, et al. Efficacy and safety of rituximab in patients with active proliferative lupus nephritis: the lupus nephritis assessment with rituximab study. *Arthritis Rheum* 2012;64(4):1215–26.
- [16] Rovin BH, van Vollenhoven RF, Aranow C, et al. A multicenter, randomized, double-blind, placebo-controlled study to evaluate the efficacy and safety of treatment with sirukumab (CNT0 136) in patients with active lupus nephritis. *Arthritis Rheumatol* 2016;68(9):2174–83.
- [17] Presto JK, Okon LG, Feng R, et al. Computerized planimetry to assess clinical responsiveness in a phase II randomized trial of topical R333 for discoid lupus erythematosus. *Br J Dermatol* 2018;178(6):1308–14.
- [18] Banchereau R, Hong S, Cantarel B, et al. Personalized immunomonitoring uncovers molecular networks that stratify lupus patients. *Cell* 2016;165(3):551–65.
- [19] Munroe ME, Vista ES, Guthridge JM, Thompson LF, Merrill JT, James JA. Proinflammatory adaptive cytokine and shed tumor necrosis factor receptor levels are elevated preceding systemic lupus erythematosus disease flare. *Arthritis Rheumatol* 2014;66(7):1888–99.
- [20] Munroe ME, Vista ES, Merrill JT, Guthridge JM, Roberts VC, James JA. Pathways of impending disease flare in African–American systemic lupus erythematosus patients. *J Autoimmun* 2017;78:70–8.
- [21] Hochberg MC. Updating the American college of rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997;40(9):1725.
- [22] Petri M, Kim MY, Kalunian KC, et al. Combined oral contraceptives in women with systemic lupus erythematosus. *N Engl J Med* 2005;353(24):2550–8.
- [23] Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing illumina microarray. *Bioinformatics* 2008;24(13):1547–8.
- [24] Chiche L, Jourde-Chiche N, Whalen E, et al. Modular transcriptional repertoire analyses of adults with systemic lupus erythematosus reveal distinct type I and type II interferon signatures. *Arthritis Rheumatol* 2014;66(6):1583–95.
- [25] Banchereau R, Cepika AM, Banchereau J, Pascual V. Understanding human autoimmunity and autoinflammation through transcriptomics. *Annu Rev Immunol* 2017;35:337–70.
- [26] Jourde-Chiche N, Whalen E, Gondouin B, et al. Modular transcriptional repertoire analyses identify a blood neutrophil signature as a candidate biomarker for lupus nephritis. *Rheumatology* 2017;56(3):477–87.
- [27] Arbuckle MR, McClain MT, Rubertone MV, et al. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *N Engl J Med* 2003;349(16):1526–33.
- [28] Bruner BF, Guthridge JM, Lu R, et al. Comparison of autoantibody specificities between traditional and bead-based assays in a large, diverse collection of patients with systemic lupus erythematosus and family members. *Arthritis Rheum* 2012;64(11):3677–86.
- [29] Dossus L, Becker S, Achaintre D, Kaaks R, Rinaldi S. Validity of multiplex-based assays for cytokine measurements in serum and plasma from “non-diseased” subjects: comparison with Elisa. *J Immunol Methods* 2009;350(1–2):125–32.
- [30] Lu R, Guthridge JM, Chen H, et al. Immunologic findings precede rapid lupus flare after transient steroid therapy. *Sci Rep* 2019;9(1):8590.
- [31] Felten R, Sagez F, Gavand PE, et al. 10 most important contemporary challenges in the management of SLE. *Lupus Sci Med* 2019;6(1):e000303.
- [32] Rai R, Chauhan SK, Singh VV, Rai M, Rai G. RNA-seq analysis reveals unique transcriptome signatures in systemic lupus erythematosus patients with distinct autoantibody specificities. *PLoS ONE* 2016;11(11):e0166312.
- [33] Ding Y, Li H, He X, et al. Identification of a gene-expression predictor for diagnosis and personalized stratification of lupus patients. *PLoS ONE* 2018;13(7):e0198325.