Stratification of Systemic Lupus Erythematosus Patients with Gene Expression Data Reveals Expression of Distinct Immune Pathways

Aditi Deokar[1]

[1]Boston University Academy, Boston, MA

August 3, 2020

Abstract

Systemic lupus erythematosus (SLE) is the tenth leading cause of death in females 15-24 years old in the US. The diversity of symptoms and immune pathways expressed in SLE patients causes difficulties in treating SLE as well as in new clinical trials. This study used unsupervised learning on gene expression data from adult SLE patients to separate patients into clusters. The dimensionality of the gene expression data was reduced by three separate methods (PCA, UMAP, and a simple linear autoencoder) and the results from each of these methods were used to separate patients into six clusters with k-means clustering. The clusters revealed three separate immune pathways in the SLE patients that caused SLE. These pathways were: (1) high interferon levels, (2) high autoantibody levels, and (3) dysregulation of the mitochondrial apoptosis pathway. Mitochondrial apoptosis has not been investigated before to our knowledge as a standalone cause of SLE, independent of autoantibody production, and mitochondrial proteins could be investigated as a therapeutic target for SLE in the future.

1 Introduction

Systemic lupus erythematosus (SLE) is the tenth most common cause of death among females 15-24 years old in the US [23]. SLE is one of many autoimmune diseases, which are diseases in which a patient's immune system mistakes parts of their own body as foreign, attacking their body instead of protecting it. In SLE specifically, the immune system attacks healthy organs and tissue. Patients can manifest very different symptoms including fatigue, swelling in the joints, the characteristic lupus butterfly-shaped rash, sunlight sensitivity, and many more [14].

SLE can be driven by defects in the innate immune system and/or the adaptive immune system. SLE patients are often characterized by high levels of interferon-1, which causes inflammation in the innate immune system in response to viruses. In SLE, high interferon levels can be caused by a variety of factors, including autoantibody complexes and neutrophil extracellular traps [3]. Most SLE patients also have high levels of autoantibodies, which are created by mature B cells (plasma cells) that were not eliminated by the tolerance mechanisms that usually prevent maturation of self-reactive B cells [7]. Autoantibodies cause a much more targeted adaptive immune response against specific self-antigens, but SLE patients can have a wide range of autoantibodies - one study found over 180 autoantibodies expressed in SLE patients [22]. Some patients with lupus do not even have autoantibodies, and many of these autoantibodies are also found in other rheumatic diseases [8].

2 Literature Review

The heterogeneity of lupus symptoms and immune pathways affected makes it difficult to treat, because different drugs work well on different patients. This heterogeneity also causes difficulties in clinical trials of more targeted biologic drugs. Merrill et al. [17] found that certain standard drugs (anti-rheumatic drugs and immunosuppressants) affect immune pathways differently in interferon-low and interferon-high patients. While there is still debate on whether SLE is one disease or many [1], it is clear that subdividing SLE patients into categories will help treatment of patients and clinical trials.

Previous studies have tackled this problem by dividing patients based on antibody levels [2], gene expression [20], and immune molecule levels [10]. However, none of these studies have reached a consensus on the best subdivision of SLE that holds true across multiple studies. Guthridge et al. [9] used all three of these factors to divide SLE patients into seven clusters with unsupervised machine learning. Using only Guthridge et al.'s gene expression data, we used different machine learning methods to create another set of clusters of SLE patients. These clusters were then used in comparison with Guthridge et al.'s clusters to determine if gene expression data alone reveals similar patterns in immune pathway expression as does its combination with antibody levels and immune molecule levels.

3 Purpose

1. Identify clusters in SLE patients from gene expression data collected by Guthridge et al. [9].
2. Identify molecular pathways expressed in clusters.
3. Compare to clusters made by previous studies, particularly Guthridge et al. [9].

4 Methods

We used a gene expression dataset available on GEO (accession number GSE138458) containing data collected by Guthridge et al. [9]. The data included 336 samples in total, with 24 control patients and 198 SLE patients. 108 of the SLE patients had two or more samples taken. Data pre-normalized by Guthridge et al. [9] was used, which had gone through bgAdjust background correction, vst variance stabilizing transformation, and rank invariant normalization. Six outliers were removed by Guthridge et al. in the normalized data, including one control patient and five SLE patients. While Guthridge et al. used modular co-expression scores followed by unsupervised random forest clustering, then reduced the dissimilarity matrix from the random forest clustering into three principal components with t-SNE (t-Distributed Stochastic Neighbor Embedding), and selected the first two components as input for k-means clustering, we chose to instead use three separate dimensionality reduction techniques followed by k-means clustering.

4.1 Dimensionality reduction

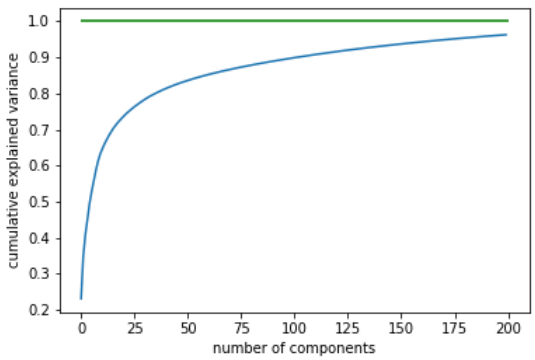


Figure 1: Cumulative variance explained by first 200 PCA components, which together explained 96.29% of variance. These 200 components were then used as features in subsequent clustering.

Prior to creating clusters from the gene expression data, the dimensionality of the 47,323 gene data was reduced using three separate methods: *Principal Component Analysis* (PCA), *Uniform Manifold Approximation and Projection* (UMAP), and a simple *autoencoder*. These techniques all reduced the 47,323 genes to fewer features in different ways to minimize the effects of random variation on the unsupervised clustering model. PCA linearly transforms data to a lower-dimensional space in a way that will maximize the variance in the data. Figure 1 depicts the cumulative explained variance for 2 to 200 principal components (the new features created by PCA that are linear combinations of the original genes). 200 principal components were selected for our model, which explain 96.29% of the variance in the original 47,323 genes.

The second technique we used, UMAP, is a nonlinear model (unlike PCA) which, similar to t-SNE, can be used for visualization, but is also used for nonlinear dimension reduction [15]. For UMAP dimension reduction, we also reduced the 47,323 genes to 200 features.

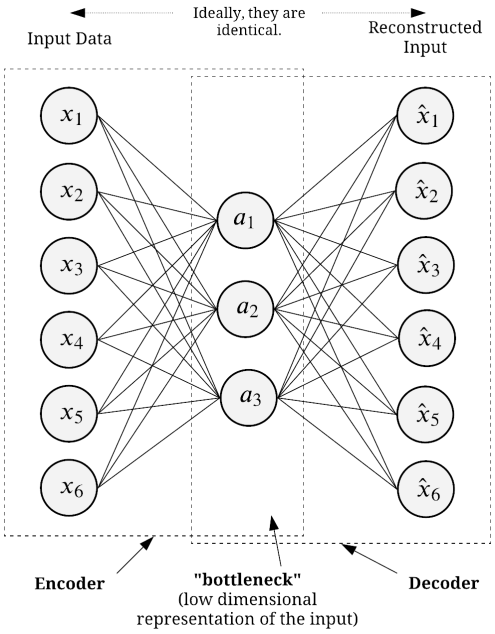
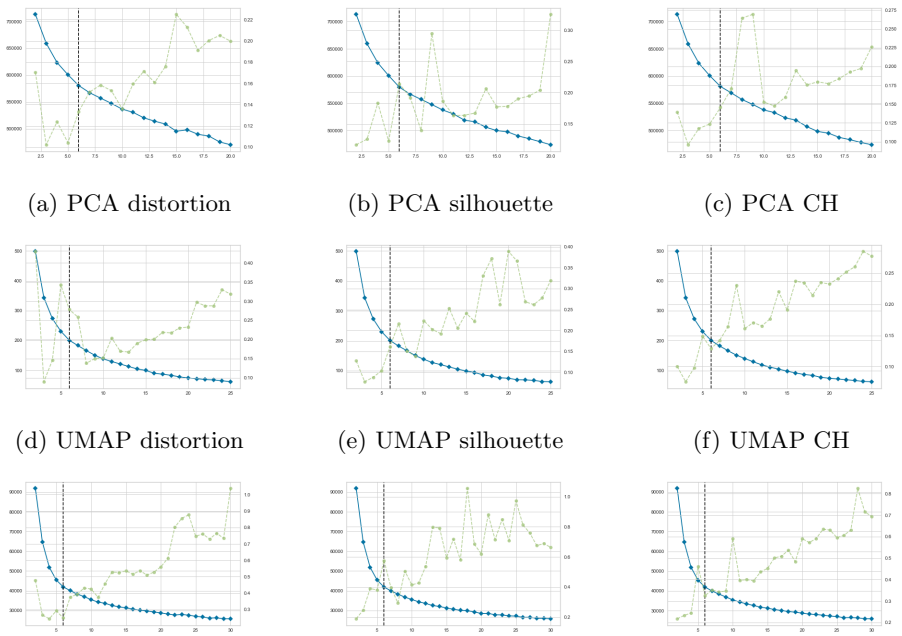


Figure 2: Diagram of simple autoencoder. Input data with a large number of features is converted to fewer features in the hidden(encoded) layer by the encoder, and then the original number of dimensions is recovered by the decoder. The loss between the original and decoded features is minimized while training. An autoencoder can be used for dimensionality reduction by selecting the encoded features.

The third technique was a simple autoencoder. Autoencoders are neural networks which, as shown in Figure 2, consist of two parts: an encoder, which reduces the dimensionality of the data, in this case from 47,323 genes to 1000 features, and a decoder, which expands these 1000 features back into 47,323 dimensions. In our study, we used the encoded data for later clustering similar

to the 200 PCA and UMAP components. A simple autoencoder, used in this study, includes only one layer in the encoder and decoder. The autoencoder aims to reduce the loss of information between the original inputs (genes) and the decoded output of the same dimension. In this case, the loss function that was optimized was the mean squared error. Another hyperparameter we tuned was the activation function. An autoencoder, like other neural networks, has an activation function which determines how the inputs of each node are converted to an output. There are many different activation functions, but two of the most common are linear and sigmoid. We trained autoencoders with both of these activations and found that the linear autoencoder performed much better after 100 epochs (validation loss 0.068) than the sigmoid autoencoder (validation loss 48.28). The linear autoencoder was then run again for 50 epochs with no change in validation loss, and the 1000 encoded components were used for subsequent clustering.

4.2 Clustering



(a) PCA distortion (b) PCA silhouette (c) PCA CH

(d) UMAP distortion (e) UMAP silhouette (f) UMAP CH

(g) AE distortion (h) AE silhouette (i) AE CH

Figure 3: Metrics used to determine optimal number of clusters for k-means clustering. Distortion scores, silhouette scores, and Calinski-Harabasz (CH) scores for data dimensionally reduced by (a-c) PCA, (d-f) UMAP, and (g-h) autoencoder (AE). All scores converged on six clusters. Graphs generated using Python YellowBrick package.

The three sets of data with reduced dimensions (200 dimensions for PCA and UMAP and 1000 dimensions for the simple linear autoencoder) were then used for k-means clustering. In order to determine the appropriate number of clusters, we used the YellowBrick Python package, which creates visualizations quantifying the "elbow method" used on the metrics distortion score (also known as Within Cluster Sum of Squares), silhouette score, and Calinski-Harabasz score (refer Figure 3). All of these metrics, for all three dimensionality reduction methods, converged on 6 clusters, which were used for k-means clustering on the three reduced dimensionality datasets.

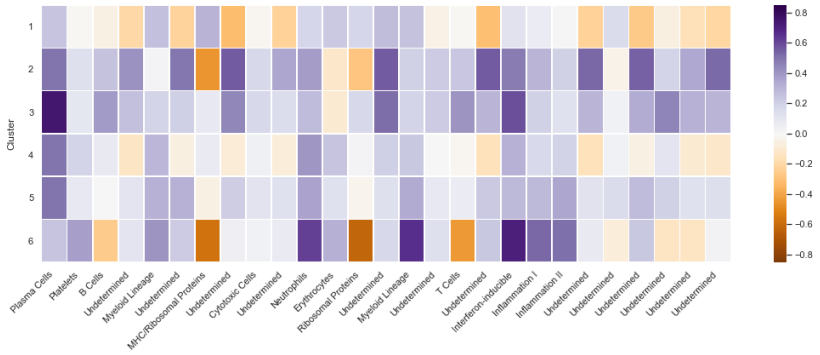
4.3 Gene modules

For visualization and interpretation of the clusters, we used 28 pre-existing modules created by Chaussabel et al. [5]. One of these modules did not contain genes in our data, so 27 modules were used to calculate module scores for the datasets from each of the three dimensionality reduction techniques (PCA, UMAP, and simple linear autoencoder). Module scores for each cluster represented the percentage of genes in each module that were significantly upregulated or downregulated by a two-tailed t-test ($p < 0.05$) in that cluster as compared to the controls.

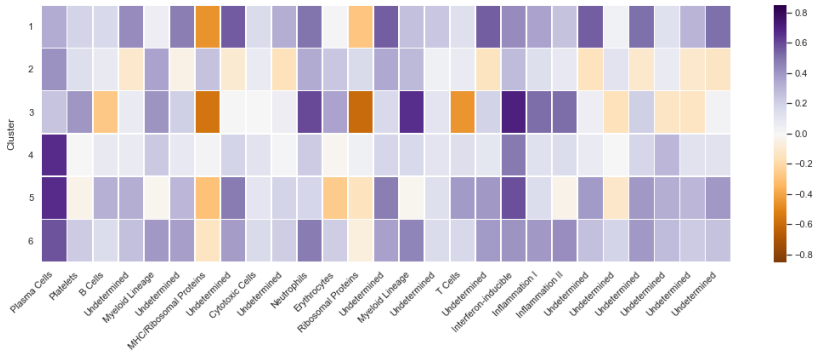
5 Results

Module scores were used to create heatmaps for data from each of the dimensionality reduction techniques (PCA, UMAP, and simple linear autoencoder) (Figure 4). Figure 5 shows 2D UMAP projections from all three dimensionality reduction techniques. In the heatmaps, the clusters originating from the PCA and UMAP dimensionality reductions showed very similar patterns in the upregulated and downregulated modules.

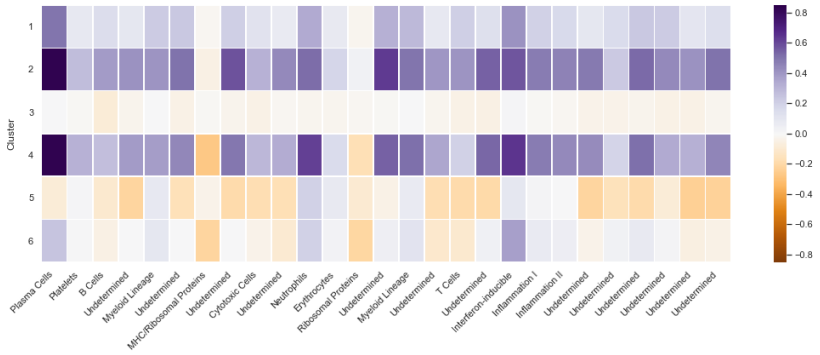
The clusters originating from the autoencoder dimensionality reduction mostly showed a consistent level of increased or decreased gene expression across all modules. Autoencoder clusters 2 and 4 showed consistent substantial upregulations across almost all genes, except for MHC class 1 proteins and ribosomal proteins (the MHC genes in the module labels MHC/Ribosomal Proteins almost entirely consist of class 1 genes as described in Chaussabel et al. [5]), which were downregulated in cluster 4 and unchanged in cluster 2. Autoencoder clusters 1 and 6 both showed only upregulation of plasma cells and interferon-inducible genes; cluster 1 additionally had some upregulation of neutrophils and cluster 6 additionally had some downregulation of MHC class 1 proteins and ribosomal proteins. Autoencoder cluster 3 had no substantial gene expression differences at all as compared to the control group in the genes present in the modules. Autoencoder cluster 5 is described below alongside PCA clusters 1 and 4 and UMAP cluster 2.



(a) PCA heatmap

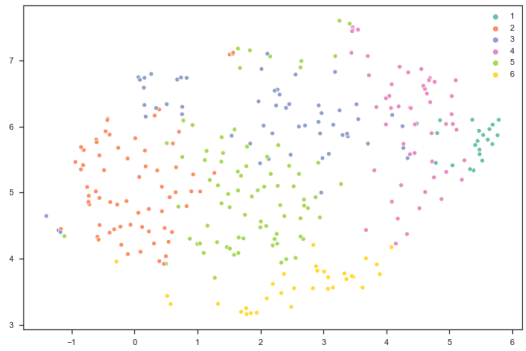


(b) UMAP heatmap

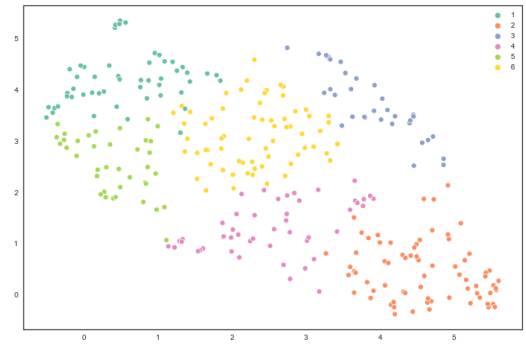


(c) Autoencoder heatmap

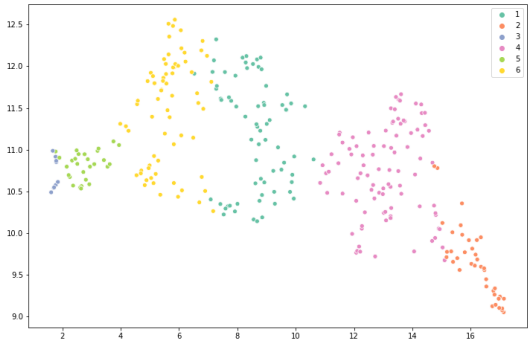
Figure 4: Heatmaps displaying the percentage of genes underexpressed (orange) or overexpressed (purple) as compared to the controls for each gene expression module (columns) in each cluster (rows). Clusters came from k-means clustering following dimensionality reduction by (a) PCA, (b) UMAP, (c) autoencoder.



(a) PCA projection



(b) UMAP projection



(c) Autoencoder projection

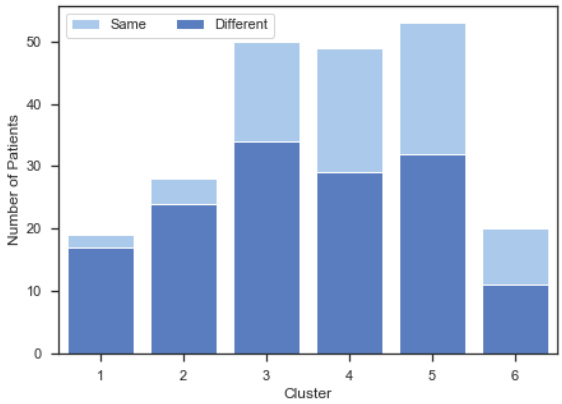
Figure 5: 2D visualizations created with UMAP of clusters created from (a) PCA, (b) UMAP, (c) autoencoder dimensionality reduction. Axes represent the first two UMAP principal components. Colors represent cluster numbers as indicated in the legend.

MHC class 1 proteins and ribosomal proteins were downregulated in PCA clusters 2 and 6 and UMAP clusters 1, 3, 5, and 6. Plasma cells were substantially upregulated in PCA clusters 2, 3, 4 and 5 and UMAP clusters 4, 5, and 6. PCA cluster 5 and UMAP cluster 4 were very similar in their upregulation of plasma cells; UMAP cluster 4 also showed some upregulation of interferon-inducible genes. Interferon-inducible genes were upregulated in PCA clusters 2, 3 and 6 and UMAP clusters 3, 4, 5, and 6.

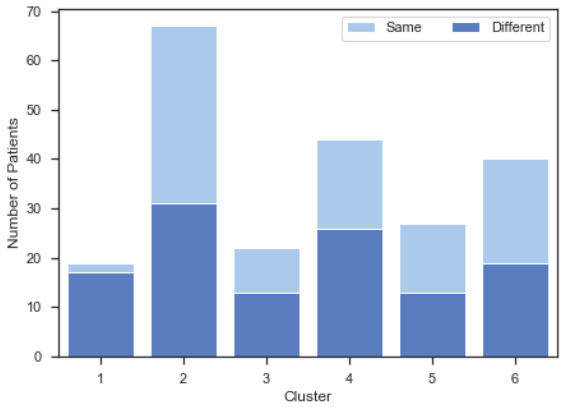
PCA cluster 6 and UMAP cluster 3 both displayed substantial upregulation of neutrophils, myeloid lineage proteins, interferon-inducible genes, and proteins related to inflammation, as well as substantial downregulation of MHC class I proteins, ribosomal proteins, B cells, and T cells.

PCA clusters 1 and 4, UMAP cluster 2, and Autoencoder cluster 5 displayed a different pattern from many of the other clusters, with many of the modules labeled as Undetermined underexpressed. These Undetermined modules represent modules which contained different types of genes and so did not have any clear uniting factor as in the other modules. The underexpressed genes in these modules (1.4, 1.6, 1.8, 2.9, 2.11, 3.4, 3.6, 3.8, 3.9) are listed in more detail in Table 1 of Chaussabel et al. [5], but included genes coding for regulators and targets of the cAMP-signaling pathway, repressors of NF-KB activation, other signaling molecules, metabolic enzymes, other enzymes, proteins involved in DNA replication, kinases and phosphatases, RAS-family members, molecules related to the cytoskeleton, T cell-expressed genes, mitochondrial ribosomal proteins, and mitochondrial elongation factors. Autoencoder cluster 5 patients additionally underexpressed module 2.1 genes (Cytotoxic cells), and genes from modules 2.7 and 3.5 (Undetermined, primarily unknown transcripts).

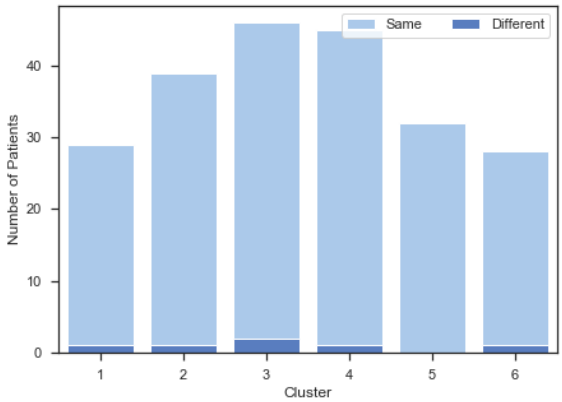
Since 108 of the SLE patients had two or more samples taken, so different samples from the same patient were included multiple times in clustering, we graphed in Figure 6 whether each of these 108 patients' samples were placed in the same cluster or different clusters for all of their samples. Each patient is counted once for every sample they provided, so a patient who was placed in clusters 1 and 3 after PCA was counted as "Different" in both cluster 1 and cluster 3 in the PCA graph, and a patient who was placed in cluster 4 for both samples taken after PCA was counted as "Same" twice in cluster 4 in the PCA graph. In clustering following PCA dimensionality reduction most of the patients who had multiple samples taken were placed in different clusters, and in clustering following UMAP about half of these patients in each cluster were placed in the same cluster and half in different clusters. In clustering following dimensionality reduction with the autoencoder, however, almost all patients who had multiple samples taken were placed in the same cluster for all of their samples.



(a) Distribution of observations in PCA clusters



(b) Distribution of observations in UMAP clusters



(c) Distribution of observations in autoencoder clusters

Figure 6: Patients in each cluster, of those who had multiple samples taken, who were put in the same cluster or different cluster for (a) PCA clusters, (b) UMAP clusters, (c) Autoencoder clusters.

6 Discussion

The aim of this study was to use unsupervised machine learning techniques on gene expression data from SLE patients to stratify patients into clusters representative of their active molecular pathways. These clusters could then be used in future clinical practice to identify the immune pathways responsible for a patient's SLE disease and prescribe treatment accordingly. We identified three sets of six clusters from gene expression data of adult SLE patients. These sets of clusters followed from three separate dimensionality reduction techniques used prior to clustering: PCA, UMAP, and a simple linear autoencoder.

For patients who had multiple samples taken, k-means following the autoencoder classified them into the same cluster 97.3% of the time, while k-means following PCA and UMAP classified them into the same cluster 32.9% and 45.7% of the time respectively (Figure 6). While gene expression data is correlated with SLE disease activity [11, 20], Petri et al. found that the majority of gene expression signatures were stable in patients over time [19]. This suggests that the autoencoder's dimensionality reduction may have emphasized the stable gene expression signatures, causing them to be a major factor in the clustering, but that PCA and UMAP, which aimed to preserve more of the variance in the data, did not maintain the data from genes whose expression was stable over time. Many of these more stable genes might not have been related to the immune system, so they were not included in Chaussabel et al.'s coexpression modules [5]. Thus, the more variable modules that were in the heatmap would have shown a lot of variation between the patients in each cluster, causing the clusters in the autoencoder heatmap to show a more consistent level of expression across all genes in the modules. Further analysis should be done to determine the level of variation in gene expression in the modules for the autoencoder clusters in comparison to the PCA and UMAP clusters, and to determine whether the PCA and UMAP clusters correlated to disease activity more than the autoencoder clusters, which this idea would imply.

The patients in the clusters created from the PCA and UMAP dimensionality reduction techniques and Autoencoder cluster 5 can be designated as belonging to one of three groups: interferon-driven SLE, antibody-driven SLE, and SLE caused by mitochondrial apoptosis.

6.1 Interferon-driven SLE

PCA cluster 6 and UMAP cluster 3 in Figure 4 both displayed substantial upregulation of interferon-inducible genes, proteins related to inflammation, neutrophils, and myeloid lineage proteins and downregulation of T cells and B cells. These two clusters correlate to Guthridge et al.'s clusters 1, 4 and 6, which also displayed these same patterns [9]. All of the upregulated genes are related to the innate immune response. In lupus, type 1 interferon levels are often elevated, which can lead to inflammation and tissue damage [6]. This pathway caused by elevated interferon levels is most likely the main cause of SLE in the patients in PCA cluster 6 and UMAP cluster 3 because of several reasons. Those

patients' have overexpressed interferon-inducible genes, and also overexpressed myeloid lineage proteins (myeloid lineage cells include pDCs and neutrophils, both of which can cause interferon levels to increase). They also have underexpressed B and T cells and normal expression of plasma cells, which would all be overexpressed if dysregulation of self-reactive B cells was the main reason for autoimmunity, rather than interferon levels.

6.2 Antibody-driven SLE

Many of the other PCA and UMAP clusters displayed upregulation of plasma cells, indicative of increased antibody production; particularly PCA clusters 2, 3, 4 and 5 and UMAP clusters 4, 5, and 6. Guthridge et al. [9] observed a similar trend, where their clusters 2, 3, and 5 had higher T cell, B cell, and plasma cell related expression. While autoantibodies are known to be common in SLE, the diversity of autoantibodies (as discussed in [22]) means that attributing the SLE of the patients in these clusters to autoantibodies in general is not enough, and there is still work to be done understanding what is different among the four PCA clusters and three UMAP clusters. Some of these differences might come from the genes used to create the clusters that were not included in the modules used for the heatmap.

Brant et al. [4], who grouped lupus patients based on their correlation between gene expression and disease activity, found three clusters, one where neutrophil levels correlated to disease activity, one where lymphocyte levels correlated to disease activity, and one more heterogeneous group. Since neutrophil extracellular traps are one way that interferon levels become elevated, their neutrophil-correlated group might correspond to our high-interferon group, and their lymphocyte-correlated group might correspond to our antibody-driven group. Once again, however, more analysis needs to be done on disease activity correlation in our data to confirm this.

6.3 SLE caused by mitochondrial apoptosis

As noted in the Results, PCA clusters 1 and 4, UMAP cluster 2, and Autoencoder cluster 5 displayed a different pattern from many of the other clusters. In these clusters, many of the modules labeled as Undetermined were underexpressed. A closer look at the genes in these Undetermined modules reveals that they included mitochondrial ribosomal proteins, mitochondrial elongation factors, and proteins in the cAMP-signaling pathway. Mitochondrial ribosomal proteins (MRPS/MRPL), in addition to their ribosomal functions, are involved in apoptotic pathways [12], and cAMP signaling regulates mitochondrial apoptosis [21]. Apoptosis is known to be a factor in SLE, but mainly because ineffective clearance of apoptotic cells can expose B and T cells to intracellular material, leading to the creation of autoantibodies against this intracellular material. SLE patients also tend to have higher rates of apoptosis generally [18].

We suggest that for the patients in PCA clusters 1 and 4, UMAP cluster 2,

and Autoencoder cluster 5, dysregulation of mitochondrial pathways or signaling from outside molecules (possibly lymphocytes) could cause mitochondrial apoptotic pathways to become activated in healthy cells, destroying healthy cells as is characteristic of SLE. These healthy cells would have a range of gene expression of mitochondrial proteins, including MRPS/MRPL, and the cells with higher expression of the proteins would activate the apoptotic pathway. Only cells with lower expression levels would survive, so lower expression levels were found in our study. These lower expression levels would also impair mitochondrial functions, which has been observed to be true in SLE patient [13]. Guthridge et al.'s cluster 7 also had low expression of mitochondrial respiration and mitochondrial stress/proteasome genes (which were not discussed in their study) [9]. The discovery of this cluster of patients using two completely different machine learning approaches suggests corroborates the idea that the mitochondrial apoptotic pathway is a novel cause for SLE.

6.4 A note on MHC class 1 and ribosomal proteins

In most of the clusters, MHC class 1 proteins and ribosomal proteins were downregulated. Both of these proteins are important in autoimmunity. MHC class 1 proteins are usually produced by all nucleated, healthy cells, which use them to display self-antigens that indicate they are healthy. This is a regulatory measure that is meant to prevent autoreactive T cells from recognizing healthy cells, and the underexpression of MHC class 1 proteins blocks this regulation [16].

Ribosomal proteins, in addition to their transcriptional role also have a role in the innate immune system. Certain ribosomal proteins, namely RPL13A and RPS3, regulate pathways that mediate inflammation [24]. Their underexpression in SLE patients would thus allow for uncontrolled inflammation.

7 Conclusion

In this study, we separated SLE patients into clusters based on their gene expression data using unsupervised learning. The data was collected by Guthridge et al. [9], who clustered patients using antibody levels and immune phenotyping in addition to gene expression levels. We used only gene expression data dn used entirely different methods from their study, in order to determine whether we would find similar clusters of patients. The dimensionality of the gene expression data was first reduced by three separate methods (PCA, UMAP, and a simple linear autoencoder) and the results from each of these methods were used to separate patients into six clusters with k-means clustering. The clusters revealed three separate immune pathways in the SLE patients that caused SLE. These pathways were 1) high interferon levels, 2) high autoantibody levels, and 3) dysregulation of the mitochondrial apoptosis pathway. All three of these pathways were present in Guthridge et al.'s clusters [9], but to our knowledge this study is the first to propose mitochondrial apoptosis as a standalone cause of SLE, independent of autoantibody production. Future studies should inves-

tigate to a further extent the mitochondrial apoptotic pathway in SLE patients as a reason for destruction of self cells in addition to a way that autoantibodies are produced.

References

[1] N. Agmon-Levin, M. Mosca, M. Petri, and Y. Shoenfeld. Systemic lupus erythematosus one disease or many? *Autoimmunity Reviews*, 11(8):593–595, 2012.

[2] Bahar Artim-Esen, Erhan Çene, Yasemin Şahinkaya, Semra Ertan, Özlem Pehlivan, Sevil Kamali, Ahmet Gül, Lale Öcal, Orhan Aral, and Murat Inanç. Cluster analysis of autoantibodies in 852 patients with systemic lupus erythematosus from a single center. *Journal of Rheumatology*, 41(7):1304–1310, 2014.

[3] Anders A. Bengtsson and Lars Rönnblom. Role of interferons in SLE. *Best Practice and Research: Clinical Rheumatology*, 31(3):415–428, 2017.

[4] Elizabeth J. Brant, Edward A. Rietman, Giannoula Lakka Klement, Marco Cavaglia, and Jack A. Tuszynski. Personalized therapy design for systemic lupus erythematosus based on the analysis of protein-protein interaction networks. *PLoS ONE*, 15(3):1–16, 2020.

[5] Damien Chaussabel, Charles Quinn, Jing Shen, Pinakeen Patel, Casey Glaser, Nicole Baldwin, Dorothee Stichweh, Derek Blankenship, Lei Li, Indira Munagala, Lynda Bennett, Florence Allantaz, Asuncion Mejias, Monica Ardura, Ellen Kaizer, Laurence Monnet, Windy Allman, Henry Randall, Diane Johnson, Aimee Lanier, Marilynn Punaro, Knut M. Wittkowski, Perrin White, Joseph Fay, Goran Klintmalm, Octavio Ramilo, A. Karolina Palucka, Jacques Banchereau, and Virginia Pascual. A Modular Analysis Framework for Blood Genomics Studies: Application to Systemic Lupus Erythematosus. *Immunity*, 29(1):150–164, 2008.

[6] Mary K. Crow. Type I Interferon in the Pathogenesis of Lupus. *The Journal of Immunology*, 192(12):5459–5468, 2014.

[7] Barbara Dema and Nicolas Charles. Autoantibodies in SLE: Specificities, Isotypes and Receptors. *Antibodies*, 5(1):2, 2016.

[8] William Egner. The use of laboratory tests in the diagnosis of SLE. *Journal of Clinical Pathology*, 53(6):424–432, 2000.

[9] Joel M. Guthridge, Rufei Lu, Ly Thi Hai Tran, Cristina Arriens, Teresa Aberle, Stan Kamp, Melissa E. Munroe, Nicolas Dominguez, Timothy Gross, Wade DeJager, Susan R. Macwana, Rebecka L. Bourn, Stephen Apel, Aikaterini Thanou, Hua Chen, Eliza F. Chakravarty, Joan T. Merrill, and Judith A. James. Adults with systemic lupus exhibit distinct molecular phenotypes in a cross-sectional study. *EClinicalMedicine*, 20:100291, 2020.

[10] Jennie A. Hamilton, Qi Wu, PingAr Yang, Bao Luo, Shanrun Liu, Jun Li, Alexa L. Mattheyses, Ignacio Sanz, W. Winn Chatham, Hui-Chen Hsu, and John D. Mountz. Cutting Edge: Intracellular IFN-β and Distinct Type I IFN Expression Patterns in Circulating Systemic Lupus Erythematosus B Cells. *The Journal of Immunology*, 201(8):2203–2208, 2018.

[11] Brian Kegerreis, Michelle D. Catalina, Prathyusha Bachali, Nicholas S. Geraci, Adam C. Labonte, Chen Zeng, Nathaniel Stearrett, Keith A. Crandall, Peter E. Lipsky, and Amrie C. Grammer. Machine learning approaches to predict lupus disease activity from gene expression data. *Scientific Reports*, 9(1):1–12, 2019.

[12] Hyun-Jung Kim, Priyanka Maiti, and Antoni Barrientos. Mitochondrial ribosomes in cancer. *Seminars in Cancer Biology*, 47(3):67–81, dec 2017.

[13] B. D. Leishangthem, A. Sharma, and Archana Bhatnagar. Role of altered mitochondria functions in the pathogenesis of systemic lupus erythematosus. *Lupus*, 25(3):272–281, 2016.

[14] Lupus Foundation of America. What is lupus?, 2020.

[15] Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. 2018.

[16] Caroline G. McPhee, Thomas J. Sproule, Dong-Mi Shin, Jason A. Bubier, William H. Schott, Martin P. Steinbuck, Lia Avenesyan, Herbert C. Morse, and Derry C. Roopenian. MHC Class I Family Proteins Retard Systemic Lupus Erythematosus Autoimmunity and B Cell Lymphomagenesis. *The Journal of Immunology*, 187(9):4695–4704, 2011.

[17] Joan T. Merrill, Fred Immermann, Maryann Whitley, Tianhui Zhou, Andrew Hill, Margot O'Toole, Padmalatha Reddy, Marek Honczarenko, Aikaterini Thanou, Joe Rawdon, Joel M. Guthridge, Judith A. James, and Sudhakar Sridharan. The Biomarkers of Lupus Disease Study: A Bold Approach May Mitigate Interference of Background Immunosuppressants in Clinical Trials. *Arthritis and Rheumatology*, 69(6):1257–1266, 2017.

[18] Dror Mevorach. Systemic Lupus Erythematosus and Apoptosis. *Clinical reviews in allergy & immunology*, 25:49–59, 2003.

[19] Michelle Petri, Wei Fu, Ann Ranger, Norm Allaire, Patrick Cullen, Laurence S. Magder, and Yuji Zhang. Association between changes in gene signatures expression and disease activity among patients with systemic lupus erythematosus. *BMC Medical Genomics*, 12(1):1–9, 2019.

[20] Daniel Toro-Domínguez, Jordi Martorell-Marugán, Daniel Goldman, Michelle Petri, Pedro Carmona-Sáez, and Marta E. Alarcón-Riquelme. Stratification of Systemic Lupus Erythematosus Patients Into Three Groups of Disease Activity Progression According to Longitudinal Gene Expression. *Arthritis and Rheumatology*, 70(12):2025–2035, 2018.

[21] Federica Valsecchi, Lavoisier S. Ramos-Espiritu, Jochen Buck, Lonny R. Levin, and Giovanni Manfredi. cAMP and mitochondria. *Physiology*, 28(3):199–209, 2013.

[22] Gal Yaniv, Gilad Twig, Dana Ben Ami Shor, Ariel Furer, Yaniv Sherer, Oshry Mozes, Orna Komisar, Einat Slonimsky, Eyal Klang, Eyal Lotan, Mike Welt, Ibrahim Marai, Avi Shina, Howard Amital, and Yehuda Shoenfeld. A volcanic explosion of autoantibodies in systemic lupus erythematosus: A diversity of 180 different antibodies found in SLE patients. *Autoimmunity Reviews*, 14(1):75–79, 2015.

[23] Eric Y Yen and Ram R Singh. Lupus – An Unrecognized Leading Cause of Death in Young Women: Population-based Study Using Nationwide Death Certificates, 2000–2015. *Arthritis and Rheumatology*, 70(8):1251–1255, 2018.

[24] Xiang Zhou, Wen Juan Liao, Jun Ming Liao, Peng Liao, and Hua Lu. Ribosomal proteins: Functions beyond the ribosome. *Journal of Molecular Cell Biology*, 7(2):92–104, 2015.