Supplementary Materials for

Machine learning reveals distinct gene signature profiles in lesional and nonlesional regions of inflammatory skin diseases

Brittany A. Martinez*, Sneha Shrotri, Kathryn M. Kingsmore, Prathyusha Bachali, Amrie C. Grammer, Peter E. Lipsky*

*Corresponding author. Email: brittany.martinez@ampelbiosolutions.com (B.A.M.); peterlipsky@comcast.net (P.E.L.)

Published 29 April 2022, Sci. Adv. 8, eabn4776 (2022)
DOI: 10.1126/sciadv.abn4776

This PDF file includes:

Supplementary Methods
Figs. S1 to S26
Tables S1 to S4
References
Supplementary Methods

Raw data processing:

*Microarray data:* Microarray data was normalized using either GeneChip Robust Multiarray Average (GCRMA), Robust Multiarray Average (RMA), or normexp background correction (NEQC) based on the microarray platform. Outliers and batch effects were identified using principal component analysis (PCA) plots. For the dataset with known batch effects, GSE81071, raw gene expression values were normalized using 11 housekeeping genes, which were shown to not vary significantly across datasets (71). These 11 housekeeping genes were: chromosome 1 Open Reading Frame 43 (C1orf43), Charged multivesicular body protein 2A (CHMP2A), ER membrane protein complex subunit 7 (EMC7), glucose-6-phosphate isomerase (GPI), proteosome subunit beta type 2 (PSMB2), proteosome subunit beta type 4 (PSMB4), member RAS oncogene family (RAB7A), receptor accessory protein 5 (REEP5), small nuclear ribonucleoprotein D3 (SNRPD3), valosin containing protein (VCP), and vacuolar protein sorting 29 homolog (VPS29).

*RNAseq data:* SRA toolkit (NCBI Sequence Read Archive, Version 2.10) was used to fetch .sra files from GEO and convert them to .fastq files. Quality of the FASTQ files was checked using FASTQC software (Babraham Institute Bioinformatics, Version 0.11.9). Adapters were removed using Trimmomatic software (Version 0.4) and appropriate head crop parameters. Trimmed reads were aligned to the human reference genome (hg38) using STAR aligner (Version 2.7). STAR output .sam files were converted to .bam files using sambamba (Version 0.8). Read summarization was provided using the featureCounts function of the Subread (Version 2.0) package. Count normalization and regularized log transformation were carried out using rlog function in DESeq2 (Version 1.32) R package.
**Gene Set Variation Analysis:** Gene Set Variation Analysis (GSVA) is a non-parametric, unsupervised method for estimating variations in gene set enrichment among the samples of an expression dataset. The GSVA algorithm was implemented using the R Bioconductor open-source package gsva (version 1.40). GSVA was carried out in one of the following ways:

*When individual datasets were analyzed*, the preprocessed log2 gene expression matrix of each dataset was used as the GSVA input. GSVA was run on each dataset separately. Before running GSVA, input genes were filtered and only those with interquartile range (IQR) of expression > 0 across all the samples were considered for analysis. All analysis in Figs. 1-6 are the result of this GSVA process. A minimum of 2 genes was required for each signature.

*For the analysis of pooled nonlesional and control samples*, log2 gene expression values generated from independent preprocessing of all 16 datasets were concatenated to create a matrix whose rows consisted of 8425 genes detected across all datasets and whose columns consisted of the 1065 samples comprised of DLE, nonlesional DLE, ACLE, SCLE, PSO, nonlesional PSO, AD, nonlesional AD, SSc, and CTLs. Log2 values were then transformed to Z-scores using `scale()` function in R. Z-score transformation converts each sample to have expression values with mean and unit variance equal to 0 (73, 74). This transformation permitted comparison of nonlesional disease samples to control directly. GSVA was then run on the following three inputs 1) 21 pooled nonlesional DLE and 168 pooled control samples, 2) 132 pooled nonlesional AD and 168 pooled control samples, and 3) 163 pooled nonlesional PSO and 168 pooled samples. The data presented in Fig. 7A is derived from this GSVA process. A minimum of 2 genes was required for each signature.

**GSVA Gene Sets:** The gene sets used for GSVA can be found in tables S2A-D.
Cellular / pathway signatures: Gene sets employed in our GSVA analysis included 48 annotated and novel cellular and pathway signatures that have been implicated in lupus (4, 5, 6) or inflammatory skin diseases (7, 8). Immune cell gene sets were previously evaluated (21, 27) or amended slightly based upon data from the Human Protein Atlas (75). Non-hematopoietic cell signatures were derived from the Human Protein Atlas (75), previously published gene sets (76), and literature mining as previously described and employed (table S2A). Pathway gene signatures were previously evaluated in lupus (21, 27), previously published (23, 24), or newly adopted by literature mining (table S2B). The output GSVA scores of each signature were used as features for training and validating ML classifiers. 40 of the 48 cellular and pathway gene signatures were used to implement the GSVA analysis on pooled nonlesional and control samples. The following signatures were excluded from the pooled nonlesional GSVA analysis because of insufficient gene numbers (<= 2) in the 8,425 genes used: LDG, GC B cell, erythrocyte, IL1 cytokines, IL12 complex, IL21 complex, IL23 complex, and the immunoproteasome.

Keratinocyte signatures: 30 gene sets specific to keratinocytes treated with individual cytokines were created from previously published studies. Only those genes that are upregulated in keratinocytes when treated with various cytokines were included in these sets (table S2C).

T cell signatures: Gene sets for T cells were created from literature mining and the Human Protein Atlas (75) to distinguish seven different T cell subsets that have been implicated in inflammatory skin disease (table S2D).

Classification and Regression Tree (CART): The library rpart (Version 4.1) was used to implement the CART algorithm for classification described previously (77, 78) and library
rpart.plot (Version 3.1) was used to visualize classification trees. GSVA enrichment scores of cellular and pathway signatures were used as independent variables and specific lesional disease (either DLE, PSO, AD, SSC, or CTL) was used as the dependent variable for analysis. Classification trees were built independently for each disease.

ML Analysis:

Creating input for ML: The input for ML was created by pooling GSVA enrichment scores of cellular and pathway gene signatures from multiple skin datasets based on sample properties (table S3B). For every dataset, GSVA enrichment scores, that range from -1 to +1, were concatenated from multiple datasets, providing a sufficiently large cohort to train and validate various ML algorithms. 14 input data frames were created for 14 separate binary ML classifications (table S3A). Seven of the 14 binary classifications involved comparing control samples (164 CTL) with either lesional samples (DLE, PSO, AD or SSC) or nonlesional samples (DLE, PSO or AD) of inflammatory skin diseases (table S3A A-D and I-K), whereas the other six binary classifications involved comparing lesional DLE samples with lesional samples of other diseases (either PSO, AD or SSC) (table S3A E-H) and nonlesional DLE samples with nonlesional samples of other diseases (table S3A L-M). In addition, another binary classification compared nonlesional PSO and nonlesional AD (table S3A). For lesional skin classification, pooled samples resulted in 90 DLE, 132 AD, 97 SSC, and 183 PSO samples. For nonlesional skin classification, pooled samples resulted in 21 DLE, 163 PSO, and 132 AD samples, and for healthy skin pooled samples were 164 CTL (table S3B).

Class balance strategies: Four class balance strategies, including: random undersampling (table S3A C), random oversampling (table S3A E, K) removing samples from an entire dataset (table S3A F), and Synthetic Minority Oversampling Technique (SMOTE) (79) (table S3A I, L, M) were used for classifications with class imbalance. The random undersampling strategy
involves randomly selecting samples from the majority class, whereas the random oversampling strategy involves randomly duplicating examples from the minority class. SMOTE functions by randomly selecting samples from the minority class, finding its k nearest neighbors, randomly selecting a neighbor, and generating a synthetic sample at a randomly selected point between two samples in the feature space. As previously noted, we used random undersampling to trim the number of examples in the majority class then used SMOTE to oversample the minority class to balance class distribution. The purpose of all class balancing strategies was to have balanced representation of both classes for ML. The dataset was split into 70% training and 30% validation and class balancing strategies were applied on the training dataset. ML algorithms were then implemented, and evaluation matrices were noted. Receiver Operating Characteristic (ROC) curves and Precision-Recall (PR) curves were plotted using the matplotlib (Version 3.3.4) library of Python. A ROC curve is graphical way to visualize trade-off between sensitivity and specificity. High area under the curve represents a low false-positive rate and a high true-positive rate. A PR curve is a measure of classification when classes are imbalanced. High area under the PR curve represents both high recall and high precision, where high precision relates to a low false-positive rate, and high recall relates to a low-false negative rate. For our analysis, we were interested in features that contributed the most towards separation of classes, hence RF was chosen as the primary ML classifier because it gives impurity-based feature importance. The top 15 features with decreasing Gini index from each classification were summarized in a bar graph using ggplot2 (Version 3.3.5) library in R. Capability of the top 15 features alone to separate the two respective classes was tested by repeating the 14 binary ML classifications using only the top 15 features. Various overlaps between the top 15 features of multiple classifications were visualized in Venn diagrams.

**Binary ML classification:** 14 separate binary ML classifications were carried out using scikit-learn (Version 0.24.1) library in Python (Version 3.8.2). For each binary classification,
performance of several ML algorithms, including: logistic regression (LR), k-nearest neighbor
(KNN), naïve Bayes (NB), support vector machines (SVM), random forest (RF), and gradient
boosting (GB) was evaluated based on sensitivity, specificity, Cohen’s kappa score, f-1 score,
and accuracy. RF was chosen as the primary ML classifier because it gives impurity-based
feature importance. The top 15 features with decreasing Gini index from each classification
were summarized in a bar graph using ggplots2 (Version 3.3.5) library in R. Capability of the top
15 features to separate two respective classes was tested by repeating the 14 binary ML
classifications using only the top 15 features.

Feature correlation: Before carrying out binary ML classification, feature selection was
necessary in order to remove noninformative or redundant features. We assessed feature
redundancy by calculating the Pearson correlation between each feature and every other
feature. Pearson correlation between features was computed using the cor function in R. corplot
library in R was used to plot 22 Pearson correlation plots (figs. S7,S10,S14,S16,S25). In 13 of
these correlation plots, there was a pair of highly correlated features (correlation coefficient >
0.8), and the feature with the lower correlation was removed using a greedy elimination
approach; this allowed us to retain the most informative features for ML (table S3A). Pearson
correlation plots were also plotted for keratinocytes gene signatures and T cell signatures (figs.
S6,S12). High correlation between the keratinocyte gene signatures made them unsuitable for
ML analysis (fig. S6).

Statistical Analysis: Statistical differences between cohorts were evaluated using Welch’s t-
test for lesional disease versus control GSVA scores from a single dataset, nonlesional samples
versus control GSVA scores from combined datasets, mean Z-scores of nonlesional samples
versus mean Z-scores of control samples of a single gene signature and Paired t-test for
lesional versus nonlesional comparison. The magnitude of this difference (the effect size) was estimated using Hedge’s $g$ calculated as below.

$$
g = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}} \frac{1}{n_1 + n_2 - 2}
$$

where,

\(\bar{x}_1 \) and \(\bar{x}_2\) = cohort 1 mean and cohort 2 mean respectively

\(n_1\) and \(n_2\) = cohort 1 size and cohort 2 size respectively

\(s_1^2\) and \(s_2^2\) = cohort 1 variance and cohort 2 variance respectively

cohort 1 and cohort 2 could be either disease and their respective control samples of a single dataset or nonlesional samples and control samples from combined dataset or mean z scores of nonlesional samples and mean Z-scores of control samples of a single gene signature or lesional and their paired nonlesional samples of a single dataset. All the statistical analysis was carried out in using effectSize (version 0.8.1) and stats (version 3.6.2) libraries in R.

**Data Visualization:** Heatmaps of GSVA Hedges’ $g$ effect size and violin plots of GSVA enrichment scores were visualized using GraphPad PRISM (Version 9.2.0). GSVA enrichment scores of gene signatures were visualized using violin plots in Prism or ComplexHeatmap (80) for hierarchical clustering (Version 2.8) package in R. Figures were made using Adobe Illustrator Creative Cloud (Version 25.3.1).
Supplemental Figures:

A

B

[Graphical representation of Supplemental Figures A and B, showing various biological data with different colors for CTL and DLE groups.]
Supplemental Figure 1: Analysis of cellular and molecular pathway signatures in lesional DLE shows increased expression of inflammatory pathways regulated by monocyte, B cell, T cell and plasmacytoid dendritic cell signatures. GSVA enrichment scores of (A) cellular gene signatures and (B) pathway gene signatures in five datasets including DLE samples (pink) and control samples (grey). The number of DLE samples per dataset that lie -1 standard deviation of the mean of the control samples is denoted on the first subtext line. The number of DLE samples per dataset that lie +1 standard deviation of the mean of the control samples is denoted on the second subtext line. Welch’s t-test: * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.
Supplemental Figure 2: Analysis of cellular and molecular pathway signatures in lesional PSO shows increased expression of keratinocyte cell signatures as well as TNF and Th17 pathway gene signatures. GSVA enrichment scores of (A) cellular gene signatures and (B) pathway gene signatures in three datasets including PSO samples (blue) and control samples (grey). The number of PSO samples per dataset that lie -1 standard deviation of the mean of the control samples is denoted on the first subtext line. The number of PSO samples per dataset that lie +1 standard deviation of the mean of the control samples is denoted on the second subtext line. Welch’s t-test: * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.
Supplemental Figure 3: Analysis of cellular and molecular pathway signatures in lesional AD shows increased expression of skin-specific dendritic cell, B cell and IL12 inflammatory pathway gene signatures. GSVA enrichment scores of (A) cellular gene signatures and (B) pathway gene signatures in two datasets including AD samples (yellow) and control samples (grey). The number of AD samples per dataset that lie -1 standard deviation of the mean of the control samples is denoted on the first subtext line. The number of AD samples per dataset that lie +1 standard deviation of the mean of the control samples is denoted on the second subtext line. Welch's t-test: * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.
Supplemental Figure 4: Analysis of cellular and molecular pathway signatures in lesional SSc samples shows increased expression of myeloid-specific cell and TGFβ fibroblast gene signatures. GSVA enrichment scores of (A) cellular gene signatures and (B) pathway gene signatures in three datasets including SSc samples (green) and control samples (grey). The number of SSc samples per dataset that lie -1 standard deviation of the mean of the control samples is denoted on the first subtext line. The number of SSc samples per dataset that lie +1 standard deviation of the mean of the control samples is denoted on the second subtext line. Welch’s t-test: * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.
### Disease vs. Control Lesional Results

| Classifier | Sensitivity | Specificity | Cohen’s kappa | Precision | F1-score | Accuracy |
|------------|-------------|-------------|---------------|-----------|----------|----------|
| Logistic R | 0.96        | 0.94        | 0.83          | 0.89      | 0.89     | 0.89     |
| DLE        | 0.93        | 0.95        | 0.87          | 0.91      | 0.93     | 0.93     |
| AD         | 0.84        | 0.85        | 0.84          | 0.86      | 0.89     | 0.89     |
| SVM        | 0.87        | 0.88        | 0.88          | 0.90      | 0.88     | 0.88     |

### ROC Curve

- **A**: DLE vs. CTL Lesional
- **B**: PSO vs. CTL Lesional
- **C**: AD vs. CTL Lesional
- **D**: SSc vs. CTL Lesional

### PR Curve

- **A**: DLE vs. CTL Lesional
- **B**: PSO vs. CTL Lesional
- **C**: AD vs. CTL Lesional
- **D**: SSc vs. CTL Lesional
**Supplemental Figure 5:** ML accurately classifies lesional skin and control skin samples.

ROC curve and PR curve of all ML algorithms to separate lesional samples from healthy control samples using all cellular and pathway gene signatures/features. ML classifiers include: logistic regression (LR, blue), k-nearest neighbors (KNN, orange), random forest (RF, green), naïve Bayes (NB, red), support vector machine (SVM, purple) and gradient boosting (GB, brown). (A) DLE versus control; (B) PSO versus control; (C) AD versus control; and (D) SSc versus control. (E) Classification metrics including sensitivity, specificity, Cohen's kappa score, precision, f-1 score and accuracy to properly separate lesional disease samples (DLE, PSO, AD or SSc) from healthy control samples with each ML classifier. Refer to tables S3A-B for details about ML. Collinear features were removed (fig. S6)
Supplemental Figure 6: Analysis of correlated features from cellular and pathway signatures was used to extract collinear features for lesional ML binary classifications. Correlation plots of GSVA enrichment scores of pooled control samples and pooled lesional (A) DLE, (B) PSO, (C) AD and (D) SSc samples. Black boxes indicate collinear samples with Pearson correlation coefficient greater than 0.8, then the feature with the lower correlation was removed using a greedy elimination approach.
Supplemental Figure 7: Direct comparison of DLE and PSO samples using GSVA shows key differences in enrichment of inflammatory cell and pathway signatures. (A) Hierarchical clustering (k=4 clusters) of GSVA enrichments scores of cellular and pathway gene signatures in two datasets that included DLE, PSO and healthy control samples. (B) Heatmap of GSVA enrichment scores of DLE compared to PSO samples in two datasets of cellular (left) and pathway (right) gene signatures. Heatmap visualization uses red (enriched signature, >0) and blue (decreased signature, <0). Welch’s t-test: * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.
### DLE vs. Disease Lesional Results

| Comparison     | Classifier | Sensitivity | Specificity | Cohen’s k | Precision | F-1 Score | Accuracy |
|----------------|------------|-------------|-------------|-----------|-----------|-----------|----------|
| DLE            | Logistic R | 0.78        | 0.91        | 0.69      | 0.59      | 0.68      | 0.83     | 0.86     |
|                |            | 0.65        | 0.77        | 0.43      | 0.71      | 0.68      | 0.72     |
|                | AD          | 0.50        | 0.78        | 0.29      | 0.54      | 0.67      |          |          |
|                | SSc         | 0.65        | 0.77        | 0.43      | 0.71      | 0.68      | 0.72     |
| PSO            | Random F    | 0.63        | 0.86        | 0.52      | 0.81      | 0.71      | 0.77     |          |
|                | AD          | 0.42        | 0.90        | 0.35      | 0.73      | 0.53      | 0.72     |          |
|                | SSc         | 0.62        | 0.84        | 0.46      | 0.76      | 0.68      | 0.74     |          |
|                | SVM         | 0.81        | 0.97        | 0.80      | 0.96      | 0.88      | 0.90     |          |
|                | AD          | 0.65        | 0.93        | 0.61      | 0.85      | 0.74      | 0.82     |          |
|                | SSc         | 0.69        | 0.90        | 0.60      | 0.86      | 0.77      | 0.81     |          |
| PSO            | Naive Bayes | 0.67        | 0.85        | 0.52      | 0.78      | 0.72      | 0.77     |          |
|                | AD          | 0.65        | 0.83        | 0.49      | 0.71      | 0.68      | 0.76     |          |
|                | SSc         | 0.62        | 0.81        | 0.43      | 0.73      | 0.67      | 0.72     |          |
| PSO            | KNN         | 0.63        | 0.94        | 0.58      | 0.89      | 0.74      | 0.80     |          |
|                | AD          | 0.69        | 0.93        | 0.64      | 0.86      | 0.77      | 0.84     |          |
|                | SSc         | 0.73        | 0.81        | 0.54      | 0.76      | 0.74      | 0.77     |          |
| PSO            | GB          | 0.70        | 0.85        | 0.56      | 0.79      | 0.74      | 0.78     |          |
|                | AD          | 0.69        | 0.85        | 0.55      | 0.75      | 0.72      | 0.79     |          |
|                | SSc         | 0.69        | 0.81        | 0.50      | 0.75      | 0.72      | 0.75     |          |
Supplemental Figure 8: ML accurately classifies lesional DLE from lesional PSO, AD and SSc. ROC curve and PR curve of all ML algorithms to separate lesional DLE from other inflammatory skin diseases using all cellular and pathway gene signatures/ features. ML classifiers include: logistic regression (LR, blue), k-nearest neighbors (KNN, orange), random forest (RF, green), naïve Bayes (NB, red), support vector machine (SVM, purple) and gradient boosting (GB, brown). (A) DLE versus PSO; (B) DLE versus AD; and (C) DLE versus SSc. (D) Classification metrics including sensitivity, specificity, Cohen’s kappa score, precision, f-1 score and accuracy to properly separate lesional DLE samples from lesional PSO, AD, and SSc samples with each ML classifier. Refer to tables S3A-B for details about ML. Collinear features were removed (fig. S9).
Supplemental Figure 9: Analysis of correlated features from cellular and pathway signatures was used to extract collinear features for lesional ML binary classifications compared to DLE. Correlation plots of GSVA enrichment scores of lesional DLE and lesional (A) PSO, (B) AD and (C) SSc samples. Correlations outlined in black were reduced to only include one feature. Black boxes indicate collinear samples with Pearson correlation coefficient greater than 0.8, then the feature with the lower correlation was removed using a greedy elimination approach.
Supplemental Figure 10: GSVA enrichment of lesional skin compared to nonlesional skin. Hedges’ g effect sizes of GSVA enrichment scores for paired lesional and nonlesional samples, including two DLE, four AD and three PSO datasets using (A) cellular gene signatures and (B) pathway gene signatures. Lesional samples were compared to their respective nonlesional paired samples in DLE, AD and PSO. Heatmap visualization uses red (enriched signature, >0) and blue (decreased signature, <0). Paired t-test: * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.
### Disease vs. Control Nonlesional Results

| Comparison   | Classifier | Sensitivity | Specificity | Cohen's kappal | Precision | F1 Score | Accuracy |
|--------------|------------|-------------|-------------|----------------|-----------|----------|----------|
| DLE          | Logistic R | 0.94        | 0.94        | 0.88           | 0.94      | 0.94     | 0.94     |
| PSO          | Logistic R | 0.72        | 0.79        | 0.49           | 0.82      | 0.77     | 0.75     |
| AD           | Logistic R | 0.63        | 0.81        | 0.44           | 0.79      | 0.70     | 0.72     |
| CTL          | Logistic R | 0.94        | 0.97        | 0.91           | 0.97      | 0.95     | 0.95     |
| DLE          | Random F   | 0.94        | 0.94        | 0.88           | 0.84      | 0.82     | 0.82     |
| PSO          | Random F   | 0.75        | 0.90        | 0.64           | 0.91      | 0.82     | 0.82     |
| AD           | Random F   | 0.75        | 0.90        | 0.64           | 0.91      | 0.82     | 0.82     |
| DLE          | SVM        | 0.94        | 0.94        | 0.85           | 0.94      | 0.92     | 0.92     |
| PSO          | SVM        | 0.75        | 0.80        | 0.68           | 0.86      | 0.84     | 0.84     |
| AD           | SVM        | 0.63        | 0.85        | 0.68           | 0.86      | 0.84     | 0.84     |
| DLE          | Naive Bayes| 0.91        | 0.94        | 0.85           | 0.94      | 0.92     | 0.92     |
| PSO          | Naive Bayes| 0.77        | 0.76        | 0.53           | 0.81      | 0.79     | 0.77     |
| AD           | Naive Bayes| 0.58        | 0.83        | 0.40           | 0.79      | 0.67     | 0.70     |
| DLE          | KNN        | 0.99        | 0.99        | 0.63           | 0.85      | 0.85     | 0.82     |
| PSO          | KNN        | 0.77        | 0.63        | 0.59           | 0.86      | 0.81     | 0.80     |
| AD           | KNN        | 0.56        | 0.91        | 0.46           | 0.88      | 0.68     | 0.73     |
| DLE          | GB         | 0.94        | 0.85        | 0.65           | 0.82      | 0.93     | 0.92     |
| PSO          | GB         | 0.68        | 0.79        | 0.81           | 0.74      | 0.74     | 0.73     |
| AD           | GB         | 0.79        | 0.85        | 0.64           | 0.85      | 0.82     | 0.82     |
Supplemental Figure 11: ML accurately classifies nonlesional skin and control skin samples. ROC curve and PR curve of all machine learning classification algorithms to separate nonlesional samples from healthy control samples using all cellular and pathway gene signatures/ features. ML classifiers include: logistic regression (LR, blue), k-nearest neighbors (KNN, orange), random forest (RF, green), naïve Bayes (NB, red), support vector machine (SVM, purple) and gradient boosting (GB, brown). (A) DLE versus control; (B) PSO versus control; and (C) AD versus control. (D) Classification metrics including sensitivity, specificity, Cohen’s kappa score, precision, f-1 score and accuracy to properly separate nonlesional disease samples (DLE, PSO or AD) from healthy control samples with each ML classifier. Refer to tables S3A-B for details about ML. Collinear features were removed (fig. S12).
Supplemental Figure 12: Analysis of correlated features from cellular and pathway signatures was used to extract collinear features for nonlesional ML binary classification. Correlation plots of GSVA enrichment scores of control samples and nonlesional (A) DLE, (B) PSO and (C) AD samples. Correlations outlined in black were reduced to only include one feature. Black boxes indicate collinear samples with Pearson correlation coefficient greater than 0.8, then the feature with the lower correlation was removed using a greedy elimination approach.
### C. DLE vs. Disease Nonlesional Results

| Comparison | Classifier | Sensitivity | Specificity | Cohen’s kappa | Precision | F-1 score | Accuracy |
|------------|------------|-------------|-------------|----------------|-----------|-----------|----------|
| DLE        | PSO        | Logistic R  | 0.97        | 0.94           | 0.91      | 0.94      | 0.95     | 0.95     |
|            | AD         |             | 0.92        | 0.94           | 0.96      | 0.94      | 0.94     | 0.94     |
|            | PSO        | Random F    | 0.88        | 0.97           | 0.94      | 0.97      | 0.93     | 0.93     |
|            | AD         |             | 0.99        | 0.97           | 0.96      | 0.99      | 0.98     | 0.98     |
|            | PSO        | SVM         | 0.96        | 0.99           | 0.96      | 0.99      | 0.98     | 0.98     |
|            | AD         |             | 0.96        | 0.96           | 0.97      | 0.97      | 0.93     | 0.94     |
|            | PSO        | Naive Bayes | 0.9         | 0.97           | 0.88      | 0.97      | 0.93     | 0.94     |
|            | AD         |             | 0.92        | 0.88           | 0.95      | 0.92      | 0.92     | 0.93     |
|            | PSO        | KNN         | 0.99        | 0.83           | 0.82      | 0.84      | 0.91     | 0.91     |
|            | AD         |             | 0.96        | 0.8           | 0.74      | 0.79      | 0.87     | 0.87     |
|            | PSO        | GB          | 0.94        | 0.91           | 0.85      | 0.89      | 0.92     | 0.92     |
|            | AD         |             | 0.88        | 0.93           | 0.81      | 0.91      | 0.89     | 0.91     |
Supplemental Figure 13: ML distinguishes nonlesional DLE from nonlesional PSO and nonlesional AD. ROC curve and PR curve of all machine learning classification algorithms to separate nonlesional DLE from other inflammatory skin diseases using all cellular and pathway gene signatures/features. ML classifiers include: logistic regression (LR, blue), k-nearest neighbors (KNN, orange), random forest (RF, green), naïve Bayes (NB, red), support vector machine (SVM, purple) and gradient boosting (GB, brown). (A) DLE versus PSO and (B) DLE versus AD. (C) Classification metrics including sensitivity, specificity, Cohen's kappa score, precision, f-1 score and accuracy to properly separate nonlesional DLE samples from nonlesional PSO and nonlesional AD samples with each ML classifier. Refer to tables S3A-B for details about ML. Collinear features were removed (fig. S15).
Supplemental Figure 14: ML less effectively classifies nonlesional PSO from nonlesional AD. (A) ROC curve and PR curve of all ML classification algorithms to separate nonlesional PSO from nonlesional AD samples using all cellular and pathway gene signatures/ features. ML classifiers include: logistic regression (LR, blue), k-nearest neighbors (KNN, orange), random forest (RF, green), naïve Bayes (NB, red), support vector machine (SVM, purple) and gradient boosting (GB, brown). (B) Top 15 features important in classifying nonlesional PSO from nonlesional AD using Gini feature importance. (C) Classification metrics including sensitivity, specificity, Cohen’s kappa score, precision, f-1 score and accuracy to properly separate nonlesional PSO samples from nonlesional AD samples with each ML classifier. (D) Correlation plots of GSVA enrichment scores of nonlesional PSO and nonlesional AD samples. Black boxes indicate collinear samples with Pearson correlation coefficient greater than 0.8, then the feature with the lower correlation was removed using a greedy elimination approach.
Nonlesional DLE and Nonlesional Disease

A  DLE and PSO

B  DLE and AD
Supplemental Figure 15: Analysis of correlated features from cellular and pathway signatures was used to extract collinear features for nonlesional ML binary classification compared to DLE. Correlation plots of GSVA enrichment scores of nonlesional DLE and (A) nonlesional PSO and (B) nonlesional AD samples. Black boxes indicate collinear samples with Pearson correlation coefficient greater than 0.8, then the feature with the lower correlation was removed using a greedy elimination approach.
Supplemental Figure 16: Analysis of cellular and molecular pathway signatures in nonlesional DLE shows upregulation of B cell, plasma cell and fatty acid metabolism gene signatures. GSVA enrichment scores using Z-score transformation of (A) cellular gene signatures and (B) pathway gene signatures in nonlesional DLE (light pink) and control samples (grey). The number of nonlesional DLE samples per dataset that lie -1 standard deviation of the mean of the control samples is denoted on the first subtext line. The number of DLE samples per dataset that lie +1 standard deviation of the mean of the control samples is denoted on the second subtext line. Welch's t-test: * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.
Supplemental Figure 17: Analysis of cellular and molecular pathway signatures in nonlesional PSO shows upregulation of innate immune cell and IL-17 gene signatures. GSVA enrichment scores using Z-score transformation of (A) cellular gene signatures and (B) pathway gene signatures in nonlesional PSO (light blue) and control samples (grey). The number of nonlesional PSO samples per dataset that lie -1 standard deviation of the mean of the control samples is denoted on the first subtext line. The number of PSO samples per dataset that lie +1 standard deviation of the mean of the control samples is denoted on the second subtext line. Welch's t-test: * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.
Supplemental Figure 18: Analysis of cellular and molecular pathway signatures in nonlesional AD shows upregulation of anti-inflammation, neutrophil, NK cell and Th17 gene signatures. GSVA enrichment scores using Z-score transformation of (A) cellular gene signatures and (B) pathway gene signatures in nonlesional AD (light yellow) and control samples (grey). The number of nonlesional AD samples per dataset that lie -1 standard deviation of the mean of the control samples is denoted on the first subtext line. The number of AD samples per dataset that lie +1 standard deviation of the mean of the control samples is denoted on the second subtext line. Welch's t-test: * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.
Supplemental Figure 19: Analysis of cellular and molecular pathway signatures in nonlesional DLE using mean of Z-score. Box plots of the mean of Z-scores of genes for each sample and gene category for (A) cellular gene signatures and (B) pathway gene signatures in nonlesional DLE (light pink) and control samples (grey). Welch's t-test: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. 
Supplemental Figure 20: Analysis of cellular and molecular pathway signatures in nonlesional PSO using mean of Z-score. Box plots of the mean of Z-scores of genes for each sample and gene category for (A) cellular gene signatures and (B) pathway gene signatures in nonlesional PSO (light blue) and control samples (grey). Welch’s t-test: * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.
**Supplemental Figure 21: Analysis of cellular and molecular pathway signatures in nonlesional AD using mean of Z-score.** Box plots of the mean of Z-scores of genes for each sample and gene category for (A) cellular gene signatures and (B) pathway gene signatures in nonlesional AD (light yellow) and control samples (grey). Welch’s t-test: * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.
Supplemental Figure 22: Cellular and pathway enrichment in SCLE is quantitatively similar to enrichment observed in DLE. (A) Hierarchical clustering (k=4 clusters) of DLE, SCLE, ACLE and healthy control samples from five lupus datasets using GSVA enrichment scores of cellular and pathway gene signatures. Hedges' g effect sizes of GSVA enrichment scores for (B) cellular gene signatures (left) and pathway gene signatures (right) in lesional SCLE and control samples in three datasets. Hedges' g effect sizes of GSVA enrichment scores for (C) cellular gene signatures (left) and pathway gene signatures (right) in lesional DLE and SCLE samples in three datasets. Heatmap visualization uses red (enriched signature, >0) and blue (decreased signature, <0). Welch's t-test: * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.
**Supplemental Figure 23: DLE and SCLE can be transcriptionally classified using ML.** (A) Hierarchical clustering (k=4) of DLE, SCLE and control samples from three lupus datasets based on GSVA scores of cellular and pathway gene signatures. (B) Correlation plot of GSVA enrichment scores of lesional DLE and lesional SCLE samples. (C) ROC curve and (D) PR curve separating DLE and SCLE using ML classifiers, including: logistic regression (LR, blue), random forest (RF, orange), support vector machine (SVM, green) and gradient boosting (GB, red). Random oversampling was used to adjust for class imbalance errors. (E) Top 15 features important in classifying DLE from SCLE using Gini feature importance. (F) Classification metrics including sensitivity, specificity, Cohen’s kappa score, precision, f-1 score and accuracy to properly separate DLE and SCLE. Refer to tables S3A-B for details about ML.
Supplemental Figure 24: Stimulated keratinocyte signatures are highly enriched in skin inflammatory diseases. Hedges’ g effect sizes of GSVA enrichment scores for lesional disease samples compared to their respective healthy control samples in five DLE, three PSO, two AD and three SSc datasets using curated keratinocyte-curated cellular signatures treated with various types of cytokines and immune molecules. Heatmap visualization uses red (enriched signature, >0) and blue (decreased signature, <0). Welch’s t-test: * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.
Supplemental Figure 25: Overabundance of correlated features from keratinocyte cell gene signatures. Correlation plot of GSVA enrichment scores to find keratinocyte gene signatures that are correlated to each other in (A) lesional DLE and control samples; (B) lesional PSO and control samples; (C) lesional AD and control samples; and (D) lesional SSc and control samples.
A. T Cell Signatures: Disease vs. Control

B. DLE and Control

C. PSO and Control

D. AD and Control

E. SSc and Control
Supplemental Figure 26: T cell subtype signatures are highly enriched in skin inflammatory diseases. GSVA enrichment scores for (A) T cell cellular signatures in lesional disease samples compared to their respective healthy control samples in five DLE, three PSO, two AD and three SSc datasets. Heatmap visualization uses red (enriched signature, >0) and blue (decreased signature, <0). Welch’s t-test: * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001. Correlation plots of GSVA enrichment scores to find T cell gene signatures that are correlated to each other in (B) DLE and control samples; (C) PSO and control samples; (D) AD and control samples; and (E) SSc and control samples.
| Sample Type | GEO ID | Platform | Design | Description |
|-------------|--------|----------|--------|-------------|
| Unpaired | GSE117239 | Human Genome U133A 2.0 | Paired and Unpaired | Clinical data on psoriasis vulgaris (PSO) patients. Includes 63L, 63NL (PSO) samples. |
| Paired | GSE117468 | Human Genome U133A 2.0 | Paired and Unpaired | Clinical data on discoid lupus erythematosus (DLE) patients. Includes 26+21 DLE, 23+20 SCLE samples. |
| Unpaired | GSE117469 | Human Genome U133A 2.0 | Paired and Unpaired | Clinical data on systemic sclerosis patients. Includes 19+21 SSc samples. |
| Unpaired | GSE117504 | Human Genome U133A 2.0 | Paired and Unpaired | Clinical data on systemic lupus erythematosus (SLE) patients. Includes 56+57 L, 56+57 NL samples. |
| Paired | GSE117506 | Human Genome U133A 2.0 | Paired and Unpaired | Clinical data on systemic lupus erythematosus (SLE) patients. Includes 56+57 L, 56+57 NL samples. |

**Supplemental Table 1: Publicly available datasets that were used in this study.**

**References**

1. Z. J. Elder, S. Assassi, W. R. Swindell, M. Wu, F. D. Tan, D. Khanna, D. E. Furst, D. P. Tashkin, R. R. Jahan-Tigh, M. D. Mayes, J. E. Gudjonsson, J. T. Chang, *Dissecting the heterogeneity of skin gene expression patterns in systemic sclerosis*, J. Allergy Clin. Immunol. **143**, 56–68 (2019).  
2. B. F. Chong, L. chiang Tseng, G. A. Hosler, N. M. Teske, S. Zhang, D. R. Karp, N. J. Olsen, C. Mohan, A subset of CD163+ macrophages displays mixed polarizations in discoid lupus skin, J. Invest. Dermatol. **145**, 145–153 (2019).  
3. J. Liu, C. C. Berthier, J. M. Kahlenberg, Enhanced Inflammasome Activity in Systemic Lupus Erythematosus Is Mediated via Type I Interferon–Induced Up-Regulation of Interferon Regulatory Factor 1, J. Immunol. **195**, 2695–2706 (2015).  
4. M. K. Sarkar, G. A. Hile, L. C. Tsoi, X. Xing, J. Liu, Y. Liang, C. C. Berthier, W. R. Swindell, M. T. Patrick, S. Shao, P. S. Tsou, R. Uppala, M. A. Beamer, A. Srivastava, S. L. Bielas, P. W. Harms, S. Getsios, J. T. Elder, *TLR-dependent model of cutaneous lupus–like inflammation*, J. Immunol. **195**, 2695–2706 (2015).  
5. P. Mande, B. Zirak, W. C. Ko, K. Taravati, K. L. Bride, T. Y. Brodeur, A. Deng, K. Dresser, Z. Jiang, R. Ettinger, K. A. Fitzgerald, M. D. Rosenblum, J. E. Harris, A. Marshak-Rothstein, Fas ligand promotes an inducible T cell subset in peripheral blood of patients with systemic lupus erythematosus, J. Allergy Clin. Immunol. **139**, 69–76 (2017).  
6. M. Suarez-Farinas, K. Li, S. Garcet, K. Hayden, A. Chiricozzi, I. Novitskaya, J. Fuentes-Duculan, M. Suarez-Farinas, K. Campbell, J. G. Krueger, *Modulation of inflammatory gene transcripts in psoriasis vulgaris: a genome-wide gene expression study in psoriatic lesions and keratinocyte cell lines*, J. Invest. Dermatol. **139**, 69–76 (2017).  
7. C. Brodmerkel, K. Li, S. Garcet, K. Hayden, A. Chiricozzi, I. Novitskaya, J. Fuentes-Duculan, M. Suarez-Farinas, K. Campbell, J. G. Krueger, *Modulation of inflammatory gene transcripts in psoriasis vulgaris: a genome-wide gene expression study in psoriatic lesions and keratinocyte cell lines*, J. Invest. Dermatol. **139**, 69–76 (2017).  
8. B. F. Chong, L. chiang Tseng, G. A. Hosler, N. M. Teske, S. Zhang, D. R. Karp, N. J. Olsen, C. Mohan, A subset of CD163+ macrophages displays mixed polarizations in discoid lupus skin, J. Invest. Dermatol. **145**, 145–153 (2019).  
9. G. D. Johnson, T. M. Reding, S. P. H. Diefenbach, J. B. S. Horowitz, *Generation of clonal human B cells for high-throughput studies of antibody responses in transplant recipients with chronic allograft nephropathy*, J. Am. Soc. Nephrol. **25**, 2523–2533 (2014).  
10. C. Brodmerkel, K. Li, S. Garcet, K. Hayden, A. Chiricozzi, I. Novitskaya, J. Fuentes-Duculan, M. Suarez-Farinas, K. Campbell, J. G. Krueger, *Modulation of inflammatory gene transcripts in psoriasis vulgaris: a genome-wide gene expression study in psoriatic lesions and keratinocyte cell lines*, J. Invest. Dermatol. **139**, 69–76 (2017).  
11. J. Liu, C. C. Berthier, J. M. Kahlenberg, Enhanced Inflammasome Activity in Systemic Lupus Erythematosus Is Mediated via Type I Interferon–Induced Up-Regulation of Interferon Regulatory Factor 1, J. Immunol. **195**, 2695–2706 (2015).  
12. E. Guttman-Yassky, R. Bissonnette, B. Ungar, M. Suárez-Fariñas, M. Ardeleanu, H. Esaki, M. Suprun, Y. Estrada, H. Xu, X. Peng, J. I. Silverberg, A. Menter, J. G. Krueger, R. Zhang, U. Chaudhry, B. Swanson, N. M. Ethnic differences in the B cell repertoire of psoriasis patients in Europe and North America, Ann. Rheum. Dis. **75**, 733–739 (2016).  
13. B. Ungar, A. B. Pavel, R. Li, G. Kimmel, J. Nia, P. Hashim, H. J. Kim, M. Chima, A. S. Vekaria, Y. Estrada, H. Xu, X. Peng, G. K. Singer, D. Baum, Y. Mansouri, M. Taliercio, E. Guttman-Yassky, Phase 2 randomized, double-blind study of IL-17 targeting with secukinumab in atopic dermatitis, J. Allergy Clin. Immunol. **139**, 145–153 (2017).  
14. L. Möbus, E. Rodriguez, I. Harder, D. Stölzl, N. Boraczynski, S. Gerdes, A. Kleinheinz, S. Abraham, A. Heratizadeh, C. Handrick, E. Haufe, T. Werfel, J. Schmitt, S. Weidinger, Atopic dermatitis displays stable and reproducible gene expression profiles in lesional and nonlesional skin, J. Allergy Clin. Immunol. **139**, 69–76 (2017).  
15. B. F. Chong, L. chiang Tseng, G. A. Hosler, N. M. Teske, S. Zhang, D. R. Karp, N. J. Olsen, C. Mohan, A subset of CD163+ macrophages displays mixed polarizations in discoid lupus skin, J. Invest. Dermatol. **145**, 145–153 (2019).  
16. S. Assassi, W. R. Swindell, M. Wu, F. D. Tan, D. Khanna, D. E. Furst, D. P. Tashkin, R. R. Jahan-Tigh, M. D. Mayes, J. E. Gudjonsson, J. T. Chang, *Dissecting the heterogeneity of skin gene expression patterns in systemic sclerosis*, J. Allergy Clin. Immunol. **143**, 56–68 (2019).
**Table S2A**: Genes within cell signatures.

| Cell Signature | Genes within cell signature |
|----------------|----------------------------|
| **B Cell**     | AICDA, BANK1, BLK, BLNK, CD19, CD22, CD79A, CLEC17A, CR2, DAPP1, DTX1, FCRL1, FCRL2, FCRL3, FCRL4, FCRL5, HLA-DOB, IGHD, IGHM, LY6D, MS4A1, PAX5, POU2AF1, SH2B2, TNFRSF13B, TNFRSF13C, VPREB1, VPREB3, ZBTB32, ZNF318 |
| **Endothelial Cell** | DLC1, ECSCR, EMCN, FLT1, KDR, LDB2, LRRRC32, MEIS2, PLAT, PTPRB, SELE, TM4SF1, TM4SF18, VWF |
| **Erythrocyte**  | BSG, GFI1B, GYPA, GYPB, GYPE, ICAM4, KEL, NFE2, RHD, SLC4A1, TRIM10, TSPD2 |
| **Fibroblast**   | 43894, ADAM33, ADAMTS6, AGTR1, ALPK2, ANGPTL2, ANKRD45, ANO2, ANPEP, ARMG9, ASPN, BDKRB2, BDNF, BMP8B, BMPCC, CD80, CD81, CEMIP, CHAC1, CHRM2, CLMP, CN1, COL14A1, COL3A1, COL5A1, COLEC10, CPXM2, CPZ, CRABP2, CXCL12, DCN, DDR2, DKK1, DMTRTA1, EGF, ELOVL2, EMLN1, FAM160A, FBLN7, FBN1, FG5, FMN2, FOXP2, FST, FSTL1, GFRAl, GLIS1, GLITB2, GRP176, GREM1, GREM2, GRIK2, GUYA12, HSD17B2, HSAP2, HSAP3, IL19, IQCD, KCMN2, KIRREL3, KRTAP1-5, KRTAP5-1, L1CAM, LAYN, LMOD1, LOXL4, LUM, LY6K, MAF4, MAF5, MARGP, MGP, MKX, MMP2, MXRA5, MXRA8, MYPN, NEXN, NUFASC, NID2, NT3, OLFM3, P3H3, P4HA3, PAMR1, PAX3, PCDOH2, PCDGHA3, PCDGHA7, PDE3B, PDEGFA, PDEGFB, PDGFA, PDGFRB, PDGRNB, PLA2R1, PLEKHA4, PLPP4, PLPPR4, PRG1, PSG5, PTGFRQ, PTRF, SEMA3A, SEMA5A, SEMA6D, SHOX, SLC16A2, SMM2, SPARC1L, SPHKAP, SSTR1, STC1, STXBP6, SUED5, SVEP1, TBX15, THBS2, TIMP2, TIMP3, TIMEM119, TIMEM130, TIMEM47, TRHDE, TRPC4, UACA, UBL4B, VAT1L, VEGFC, WNT5A, WNT5B, ZFPM2 |
| **GC B Cell**   | FCRLA, GCSAM, KLHL6, LRPMP, NUGGC, RGS13 |
| **Granulocyte** | CLC, HSH2D2, M4A2, PGRYR1P1, PRG2, SYNE1 |
| **Keratinocyte**| ABLIM, AKR1C1, ALDH8, ALOX12B, ANXAX, AQP3, ATDC, BPAG1, CA12, CCND2, CD24, CDH3, CDKN1A, CDSN, COL17A1, CST6, CTA1, DD96, DSC1, DSG1, DSP, EGFR, EVPL, Fgf2, G0S2, G1A1, GLUL, GNA15, HBP17, IFI27, ITIH3, ITGAM, ITGB4, IVL, JUNB, JUP, KLK1, KL7, KL9, KRT1, KRT4, KRT15, KRT16, KRT22, KRT5, KRT6A, LAMA3, LGALS7, LOR, NOTCH3, PPL, PROX2, PRSS11, PRSS24, S100A2, SERPINF2, SERPINE3, SPRED1, SPRED2, S100A20, SERPINF3, SERPINE6, SERPINE7, SPRED1, SPN, SPINK5, SPRR1B, TACSTD2, TFAP2A, TGM1, TP63, TUBA1, XP5 |
| **Langerhans Cell** | CD1B, CD1C, CD1E, CD207 |
| **LDG**         | AZU1, CAMP, CECAM3, CECAM4, CECAM6, CECAM8, CTSG, DEFA4, ELANE, MPO, OLFM4, RNAE3 |
| **Melanocyte**  | ASIP, CITED1D, DCT, GPNMB, GPR143, MITF, MLANA, MLPH, OCA2, PMEL, SLC24A5, SLC45A2, TYR, TYRP1 |
| **Monocyte**    | ADGRE1, C1OA, C1OB, C1OC, C2, CD14, CD300C, CD300E, CD5L, CD68, CLEC5A, CSF1R, CYBB, FOLR2, LILRA1, MARCO, MERTK, MS4A7, MSR1, SPIC |
| **Monocyte/Myeloid Cell** | ADGRE2, ADGRE3, AIF1, APOC1, BPI, BST1, C4A, C4B, C4BPA, C4BPN, C5, C8A, C8B, CD163, CD1D, CD209, CD200L, CD33, CFD, CFP, CHIT1, CLEC12A, CLEC12B, CLEC1A, CLEC4A, CLEC4D, CLEC4E, CLEC6A, CRISP3, CSF2RA, CSF2RB, CST, CTSS, F12, FCER1A, FCER1G, FCGR1A, FCGR1B, FCGR2A, FCGR2C, FLT3, GRN, IFG6F, ITGA5, LGALS12, LGALS4, LGALS9, LILRA2, LILRA5, LILRB2, LILRB6, LY6E, LY6V1, LYZ, MEFV, MMP8, MNDA, MPEG1, MS4A4A, MS4A6A, NLRP12, NLRP3, NOD2, OLRI1, OSCAR, OSM, PILRA, PRAM1, RETN, S100A12, S100A8, S100A9, SCARB1, SECTM1, SEMA4A, SERPING1, SGK1, SIGLEC1, SIGLEC10, SIGLEC14, SIGLEC5, SLC11A1, SLTR4, SMPD3B, SP1, TEK, THBD, TLR2, TLR9, TNFSF13B, TREM1, TREML4, TYROBP, VENTX, VSG41, VSTM1 |
| **Neutrophil**  | ARG1, BMX, CD177, CSF3R, DEFA1, DEFA1B, DEFA3, DEFB103A, DEBF103B, DEBF106B, DEBF136, DEBF4A, FPR2, ORL1J, PRITN3, SLCA23A, SLPI |
| **NK Cell**     | KIR2DL4, KLRC3, KLRF1, NCA1, NCR1, SH2D1B, TNFSF11, TXK |
| **pDC**         | CLEC4C, FNF, LILRA4, NR1P1, PDCSIN1, PLA2GS, PLAC8, PTCRA, SERPINF1, SLC15A4, TCF4 |
| **Plasma Cell** | CD38, CRELD2, ELL2, FKB11, IGKV4-1, IGLV2-14, ITM2C, JCHAIN, MANF, MZB1, PDA4, PRDX4, SD2FL1, SPATS2, TNFRSF17, UAP1 |
| **Platelet**    | GP1A8, GP5, GP6, GP9, PFAV1, PLEK, PPBP, SLC35D |
| **Skin-specific DC** | CLEC10A, CLEC9A, GPR31, MRC1, XCR1 |
| **T Cell**      | BCL11B, CAMK4, CD28, CD3D, CD3G, CD5, CD6, GPR171, ITK, KGCC, TESPA1, THEMIS, TRAT1 |
Supplemental Table 2B: Genes within pathway signatures.

Table S2B: Genes within pathway signatures.
### Table S2C: Genes within keratinocyte signatures.

| Signature | Genes in the signature |
|-----------|------------------------|
| S100A7, S100A8, MMP1, MMP3, SERPINB4, SERPINB3, SLC6A14, TGFA, HEPHL1 | |
| IL32, CCL8, CXCL3, CXCL8, CCL20, CXCL5, CCL17, CCL2, CXCL14, CCL5, CCL22, IFNL1, IFNB1, IFNL2, IFNL3, IL6, IL1B | |
| S100A7, S100A8, MMP1, MMP3, KLK7, KRTDAP, KRT10 | |
| CNFN, TREX2, SPRR2B, SPRR2D, SULT2B1 | |
| EREG, IL1B, IL36G, CXCL2, CXCL8, MMP1, MMP10, PTP4A1, MAT2A, CYP27B1, SCG5, PHLDA1, ARL4C | |
| SERPINB4, SERPINB3, SLC6A14, TGFA, HEPHL1 | |
| IL32, CCL8, CXCL3, CXCL8, CCL20, CXCL5, CCL17, CCL2, CXCL14, CCL5, CCL22, IFNL1, IFNB1, IFNL2, IFNL3, IL6, IL1B | |
**Supplemental Table 2D**: Genes within T cell signatures.

| T cell Signature   | Genes in the signature                      | Reference   |
|--------------------|---------------------------------------------|-------------|
| Dermal Aner/Act T Cell | CD160, CTLA4, ICOS, KLRG1, LAG3, PDCD1 |             |
| Dermal CD8 T Cell   | CD8A, CD8B                                  |             |
| Dermal Thf          | BTLA, IL21, SH2D1A                          |             |
| Dermal Th1          | CCL5, CXCR3, EOMES, IFNG, PRF1, TBX21, GZMK | (1-5)       |
| Dermal Th17         | CCR6, IL12RB1, IL17A, IL17F, IL22, IL23R, IL26, KLRB1, RORC |             |
| Dermal Th2          | GATA3, IL13, IL4, IL4R, IL5                 |             |
| Dermal Treg         | FOXP3, IL2F2, TNFRSF9                       |             |

References:

1. S. Crotty, Follicular helper CD4 T cells (TFH), *Annu Rev Immunol*. **29**, 621–663 (2011), doi: 10.1146/annurev-immunol-031210-101400.

2. T. Duhen, C. Ni, D. Campbell, Identification of a specific gene signature in human Th1/17 cells (BA13P.126), *J. Immunol.* **192**, 177.12 LP--177.12 (2014).

3. N. Kutukculer, E. Azarsiz, G. Aksu, N. E. Karaca, CD4+CD25+Foxp3+ T regulatory cells, Th1 (CCR5, IL-2, IFN-γ) and Th2 (CCR4, IL-4, IL-13) type chemokine receptors and intracellular cytokines in children with common variable immunodeficiency, *Int. J. Immunopathol. Pharmacol.* **29**, 241–251 (2016).

4. J. B. Wing, Y. Kitagawa, M. Locci, H. Hume, C. Tay, T. Morita, Y. Kidani, K. Matsuda, T. Inoue, T. Kurosaki, S. Crotty, C. Coban, N. Ohkura, S. Sakaguchi, A distinct subpopulation of CD25+ T-follicular regulatory cells localizes in the germinal centers, *Proc. Natl. Acad. Sci. U. S. A.* **114**, E6400–E6409 (2017).

5. B. Höllbacher, T. Duhen, S. Motley, M. M. Klicznik, I. K. Gratz, D. J. Campbell, Transcriptomic profiling of human effector and regulatory T cell subsets identifies predictive population signatures, *Immunohorizons* **4**, 585–596 (2021).
Table S3A: Class balance strategy used for machine learning classification.

| Skin Type | Comparison | Original Samples | ML Samples | Class balance strategy | Class | Greedy Elimination |
|-----------|------------|------------------|------------|------------------------|-------|-------------------|
| Lesional Skin | (A) DLE vs CTL | 90 DLE vs 164 CTL | 90 DLE vs 164 CTL | None | Class 0: CTL | IL12, Monocyte / Myeloid Cell |
| | (B) AD vs CTL | 132 AD vs 164 CTL | 132 AD vs 164 CTL | None | Class 0: CTL | Monocyte, IL12, Peroxisome |
| | (C) SSc vs CTL | 97 SSc vs 164 CTL | 97 SSc vs 97 CTL | Random Undersampling | Class 0: CTL | Monocyte, IL12, Complement Proteins, Peroxisome, Monocyte / Myeloid Cell |
| | (D) PSO vs CTL | 183 PSO vs 164 CTL | 183 PSO vs 164 CTL | None | Class 0: CTL | IL12, Monocyte |
| | (E) DLE vs SUB | 90 DLE vs 54 SUB | 90 DLE vs 96 SUB | Random Oversampling | Class 0: SUB | None |
| | (F) DLE vs PSO | 90 DLE vs 183 PSO | 90 DLE vs 110 PSO | Excluded GSE17468 | Class 0: PSO | None |
| | (G) DLE vs AD | 90 DLE vs 132 AD | 90 DLE vs 132 AD | None | Class 0: AD | Monocyte / Myeloid Cell, TNF |
| | (H) DLE vs SSc | 90 DLE vs 97 SSc | 90 DLE vs 97 SSc | None | Class 0: SSc | FABO, TNF |
| Nonlesional Skin | (I) DLE vs CTL | 21 DLE vs 164 CTL | 98 DLE vs 122 CTL | SMOTE over and undersample | Class 0: CTL | None |
| | (J) PSO vs CTL | 163 PSO vs 164 CTL | 163 PSO vs 164 CTL | Excluded GSE52471 and GSE109248 | Class 0: PSO | Monocyte |
| | (K) AD vs CTL | 132 AD vs 164 CTL | 164 AD vs 164 CTL | Random Oversampling | Class 0: AD | Monocyte |
| | (L) DLE vs PSO | 21 DLE vs 163 PSO | 97 DLE vs 121 PSO | SMOTE over and undersample | Class 0: PSO | Monocyte, FABO, IL12 |
| | (M) DLE vs AD | 21 DLE vs 132 AD | 79 DLE vs 98 AD | SMOTE over and undersample | Class 0: AD | Monocyte, T Cell, FABO, IL12 |
| | (N) PSO vs AD | 163 PSO vs 132 AD | 163 PSO vs 132 AD | None | Class 0: AD | Monocyte / Myeloid Cell, Complement Proteins, Peroxisome, IL12 |
### Table S3B: Number of samples pooled from each skin dataset to create input for machine learning.

| Samples | Datasets | Number of samples |
|---------|----------|-------------------|
| 90 DLE  | GSE52471 | 7                 |
|         | GSE100093 | 15               |
|         | GSE52471 | 3                 |
|         | GSE72535 | 9                 |
|         | GSE81071_A | 26               |
|         | GSE100093 | 6                 |
|         | GSE120809 | 6                 |
|         | GSE58095 | 39               |
|         | GSE130955 | 58               |

| Lesional samples | Non Lesional samples | Healthy control samples |
|------------------|----------------------|-------------------------|
| 132 AD           | 132 AD               | 164 CTL                |
|                  | GSE1135088 | 19                |
|                  | GSE1135090 | 38               |
|                  | GSE1135094 | 54               |
|                  | GSE1135121 | 21               |
|                  | GSE1135088 | 19                |
|                  | GSE1135090 | 38               |
|                  | GSE1135094 | 54               |
|                  | GSE1135121 | 21               |
|                  | GSE1135121 | 21               |
|                  | GSE1135121 | 21               |
|                  | GSE1135121 | 21               |
|                  | GSE1135121 | 21               |
|                  | GSE1135121 | 21               |

Table S3B: Number of samples pooled from each skin dataset to create input for machine learning.
Table S4: Comparison between mean of Z-score per gene signature and GSVA enrichment scores using Z-score transformation of nonlesional samples. Left indicates statistics from Welch’s t-test between NL and CTL samples. Right indicates statistics from Hedges’ g Effect size of NL compared to CTL samples.
REFERENCES AND NOTES

1. B. Tebbe, C. Orfanos, Epidemiology and socioeconomic impact of skin disease in lupus erythematosus. *Lupus* 6, 96–104 (1997).

2. M. P. Maz, J. Michelle Kahlenberg, Cutaneous and systemic connections in lupus. *Curr. Opin. Rheumatol.* 32, 583–589 (2020).

3. L. Uva, D. Miguel, C. Pinheiro, J. P. Freitas, M. Marques Gomes, P. Filipe, Cutaneous manifestations of systemic lupus erythematosus. *Autoimmune Dis.* 2012, 834291 (2012).

4. J. Wenzel, Cutaneous lupus erythematosus: New insights into pathogenesis and therapeutic strategies. *Nat. Rev. Rheumatol.* 15, 519–532 (2019).

5. S. Ribero, S. Sciascia, L. Borradori, D. Lipsker, The cutaneous spectrum of lupus erythematosus. *Clin. Rev. Allergy Immunol.* 53, 291–305 (2017).

6. P. Vashisht, K. Borghoff, J. R. O’Dell, M. Hearth-Holmes, Belimumab for the treatment of recalcitrant cutaneous lupus. *Lupus* 26, 857–864 (2017).

7. E. F. Morand, R. Furie, Y. Tanaka, I. N. Bruce, A. D. Askanase, C. Richez, S.-C. Bae, P. Z. Brohawn, L. Pineda, A. Berglind, R. Tummala, Trial of anifrolumab in active systemic lupus erythematosus. *N. Engl. J. Med.* 382, 211–221 (2020).

8. A. Menter, B. E. Strober, D. H. Kaplan, D. Kivelevitch, E. F. Prater, B. Stoff, A. W. Armstrong, C. Connor, K. M. Cordoro, D. M. R. Davis, B. E. Elewski, J. M. Gelfand, K. B. Gordon, A. B. Gottlieb, A. Kavanaugh, M. Kiselica, N. J. Korman, D. Kroshinsky, M. Lebwohl, C. L. Leonardi, J. Lichten, H. W. Lim, N. N. Mehta, A. S. Paller, S. L. Parra, A. L. Pathy, R. N. Rupani, M. Siegel, E. B. Wong, J. J. Wu, V. Hariharan, C. A. Elmets, Joint AAD-NPF guidelines of care for the management and treatment of psoriasis with biologics. *J. Am. Acad. Dermatol.* 80, 1029–1072 (2019).

9. D. Deleanu, I. Nedelea, Biological therapies for atopic dermatitis: An update. *Exp. Ther. Med.* 17, 1061–1067 (2019).
10. A. Jabbari, M. Suárez-Fariñas, J. Fuentes-Duculan, J. Gonzalez, I. Cueto, A. G. Franks, J. G. Krueger, Dominant Th1 and minimal Th17 skewing in discoid lupus revealed by transcriptomic comparison with psoriasis. *J. Invest. Dermatol.* **134**, 87–95 (2014).

11. B. F. Chong, L. C. Tseng, G. A. Hosler, N. M. Teske, S. Zhang, D. R. Karp, N. J. Olsen, C. Mohan, A subset of CD163⁺ macrophages displays mixed polarizations in discoid lupus skin. *Arthritis Res. Ther.* **17**, 324 (2015).

12. A. M. S. Barron, J. C. Mantero, J. D. Ho, B. Nazari, K. L. Horback, J. Bhawan, R. Lafyatis, C. Lam, J. L. Browning, Perivascular adventitial fibroblast specialization accompanies T Cell retention in the inflamed human dermis. *J. Immunol.* **202**, 56–68 (2019).

13. P. Mande, B. Zirak, W. C. Ko, K. Taravati, K. L. Bride, T. Y. Brodeur, A. Deng, K. Dresser, Z. Jiang, R. Ettinger, K. A. Fitzgerald, M. D. Rosenblum, J. E. Harris, A. Marshak-Rothstein, Fas ligand promotes an inducible TLR-dependent model of cutaneous lupus–like inflammation. *J. Clin. Invest.* **128**, 2966–2978 (2018).

14. J. Liu, C. C. Berthier, J. M. Kahlenberg, Enhanced inflammasome activity in systemic lupus erythematosus is mediated via type I interferon–induced up-regulation of interferon regulatory factor 1. *Arthritis Rheumatol.* **69**, 1840–1849 (2017).

15. L. C. Tsoi, G. A. Hile, C. C. Berthier, M. K. Sarkar, T. J. Reed, J. Liu, R. Uppala, M. Patrick, K. Raja, X. Xing, E. Xing, K. He, J. E. Gudjonsson, J. M. Kahlenberg, Hypersensitive IFN responses in lupus keratinocytes reveal key mechanistic determinants in cutaneous lupus. *J. Immunol.* **202**, 2121–2130 (2019).

16. V. P. Werth, D. Fiorentino, B. A. Sullivan, M. J. Boedigheimer, K. Chiu, C. Wang, G. E. Arnold, M. A. Damore, J. Bigler, A. A. Welcher, C. B. Russell, D. A. Martin, J. B. Chung, Brief Report: Pharmacodynamics, safety, and clinical efficacy of AMG 811, a human anti–interferon-γ antibody, in patients with discoid lupus erythematosus. *Arthritis Rheumatol.* **69**, 1028–1034 (2017).
17. M. D. Catalina, P. Bachali, N. S. Geraci, A. C. Grammer, P. E. Lipsky, Gene expression analysis delineates the potential roles of multiple interferons in systemic lupus erythematosus. *Commun. Biol.* **2**, 140 (2019).

18. K. E. Nograles, L. C. Zaba, E. Guttman-Yassky, J. Fuentes-Duculan, M. Suárez-Fariñas, I. Cardinale, A. Khatcherian, J. Gonzalez, K. C. Pierson, T. R. White, C. Pensabene, I. Coats, I. Novitskaya, M. A. Lowes, J. G. Krueger, Th17 cytokines interleukin (IL)-17 and IL-22 modulate distinct inflammatory and keratinocyte-response pathways. *Br. J. Dermatol.* **159**, 1092–1102 (2008).

19. Y. Asano, Systemic sclerosis. *J. Dermatol.* **45**, 128–138 (2018).

20. P. M. Brunner, E. Guttman-Yassky, D. Y. M. Leung, The immunology of atopic dermatitis and its reversibility with broad-spectrum and targeted therapies. *J. Allergy Clin. Immunol.* **139**, S65–S76 (2017).

21. K. M. Kingsmore, P. Bachali, M. D. Catalina, A. R. Daamen, S. E. Heuer, R. D. Robl, A. C. Grammer, P. E. Lipsky, Altered expression of genes controlling metabolism characterizes the tissue response to immune injury in lupus. *Sci. Rep.* **11**, 14789 (2021).

22. S. J. Waddell, S. J. Popper, K. H. Rubins, M. J. Griffiths, P. O. Brown, M. Levin, D. A. Relman, Dissecting interferon-induced transcriptional programs in human peripheral blood cells. *PLOS ONE* **5**, e9753 (2010).

23. J. L. Sargent, A. Milano, S. Bhattacharyya, J. Varga, M. K. Connolly, H. Y. Chang, M. L. Whitfield, A TGFB-responsive gene signature is associated with a subset of diffuse scleroderma with increased disease severity. *J. Invest. Dermatol.* **130**, 694–705 (2010).

24. C. L. Langrish, Y. Chen, W. M. Blumenschein, J. Mattson, B. Basham, J. D. Sedgwick, T. McClanahan, R. A. Kastelein, D. J. Cua, IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. *J. Exp. Med.* **201**, 233–240 (2005).

25. A. Blauvelt, A. Chiricozzi, The immunologic role of IL-17 in psoriasis and psoriatic arthritis pathogenesis. *Clin. Rev. Allergy Immunol.* **55**, 379–390 (2018).
26. J. E. Hawkes, B. Y. Yan, T. C. Chan, J. G. Krueger, Discovery of the IL-23/IL-17 signaling pathway and the treatment of psoriasis. *J. Immunol.* **201**, 1605–1613 (2018).

27. M. D. Catalina, P. Bachali, A. E. Yeo, N. S. Geraci, M. A. Petri, A. C. Grammer, P. E. Lipsky, Patient ancestry significantly contributes to molecular heterogeneity of systemic lupus erythematosus. *JCI Insight* **5**, e140380 (2020).

28. T. Vazquez, R. Feng, K. J. Williams, V. P. Werth, Immunological and clinical heterogeneity in cutaneous lupus erythematosus. *Br. J. Dermatol.* **185**, 481–483 (2021).

29. J. L. Zhu, L. T. Tran, M. Smith, F. Zheng, L. Cai, J. A. James, J. M. Guthridge, B. F. Chong, Modular gene analysis reveals distinct molecular signatures for subsets of patients with cutaneous lupus erythematosus. *Br. J. Dermatol.* **185**, 563–572 (2021).

30. L. C. Tsoi, E. Rodriguez, F. Degenhardt, H. Baurecht, U. Wehkamp, N. Volks, S. Szymczak, W. R. Swindell, M. K. Sarkar, K. Raja, S. Shao, M. Patrick, Y. Gao, R. Uppala, B. E. Perez White, S. Getsios, P. W. Harms, E. Maveraaki, J. T. Elder, A. Franke, J. E. Gudjonsson, S. Weidinger, Atopic dermatitis is an IL-13–Dominant disease with greater molecular heterogeneity compared to psoriasis. *J. Invest. Dermatol.* **139**, 1480–1489 (2019).

31. H. Valdimarsson, J. E. Gudjonsson, A. Johnston, H. Sigmundsdottir, H. Valdimarsson, Immunopathogenic mechanisms in psoriasis. *Clin. Exp. Immunol.* **135**, 1–8 (2004).

32. L. Pasquali, A. Srivastava, F. Meisgen, K. Das Mahapatra, P. Xia, N. Xu Landén, A. Pivarcsi, E. Sonkoly, The keratinocyte transcriptome in psoriasis: Pathways related to immune responses, cell cycle and keratinization. *Acta Derm. Venereol.* **99**, 196–205 (2019).

33. H. Jin, M. K. Oyoshi, Y. Le, T. Bianchi, S. Koduru, C. B. Mathias, L. Kumar, S. Le Bras, D. Young, M. Collins, M. J. Grusby, J. Wenzel, T. Bieber, M. Boes, L. E. Silberstein, H. C. Oettgen, R. S. Geha, IL-21R is essential for epicutaneous sensitization and allergic skin inflammation in humans and mice. *J. Clin. Invest.* **119**, 47–60 (2009).

34. F. Gong, Q. Su, Y. H. Pan, X. Huang, W. H. Shen, The emerging role of interleukin-21 in allergic diseases (Review). *Biomed. Rep.* **1**, 837–839 (2013).
35. A. P. Sappino, I. Masouye, J. H. Saurat, G. Gabbiani, Smooth muscle differentiation in scleroderma fibroblastic cells. *Am. J. Pathol.* **137**, 585–591 (1990).

36. J. D’Orazio, S. Jarrett, A. Amaro-Ortiz, T. Scott, UV radiation and the skin. *Int. J. Mol. Sci.* **14**, 12222–12248 (2013).

37. M. Laporte, P. Galand, D. Fokan, C. De Graef, M. Heenen, Apoptosis in established and healing psoriasis. *Dermatology* **200**, 314–316 (2000).

38. A. Trautmann, M. Akdis, S. Klunker, K. Blaser, C. A. Akdis, Role of apoptosis in atopic dermatitis. *Int. Arch. Allergy Immunol.* **124**, 230–232 (2001).

39. B. Franz, B. Fritzscheing, A. Riehl, N. Oberle, C. D. Klemke, J. Sykora, S. Quick, C. Stumpf, M. Hartmann, A. Enk, T. Ruzicka, P. H. Krammer, E. Suri-Payer, A. Kuhn, Low number of regulatory T cells in skin lesions of patients with cutaneous lupus erythematosus. *Arthritis Rheum.* **56**, 1910–1920 (2007).

40. R. Webb, J. T. Merrill, J. A. Kelly, A. Sestak, K. M. Kaufman, C. D. Langefeld, J. Ziegler, P. Robert, J. C. Edberg, R. Ramsey-goldman, M. Petri, J. D. Reveille, G. S. Alarcón, L. M. Vilá, M. E. Alarcón-Riquelme, J. A. James, G. S. Gilkeson, C. O. Jacob, K. L. Moser, P. M. Gaffney, T. J. Vyse, S. K. Nath, P. Lipsky, J. B. Harley, A. H. Sawalha, A polymorphism within interleukin-21 receptor (IL21R) confers risk for systemic lupus erythematosus. *Arthritis Rheumatol.* **60**, 2402–2407 (2009).

41. A. Pușcaș, A. Cătană, C. Pușcaș, I. Roman, C. Vornicescu, M. Șomlea, R. Orăsan, Psoriasis: Association of interleukin-17 gene polymorphisms with severity and response to treatment (Review). *Exp. Ther. Med.* **875**–**880** (2019).

42. A. C. Allison Billi, F. Ma, O. Plazyo, M. Gharaeef-Kermi, R. Wasikowski, G. A. Hile, X. Xing, C. M. Yee, S. M. Rizvi, M. P. Maz, F. Wen, L. C. Tsoi, M. Pellegrini, R. L. Modlin, J. E. Gudjonsson, J. M. Kahlenberg, A. C. Billi, F. Ma, O. Plazyo, M. G.-Kermi, R. Wasikowski, G. A. Hile, X. Xing, C. M. Yee, S. M. Rizvi, M. P. Maz, F. Wen, L. C. Tsoi, M. Pellegrini, R. L. Modlin, J. E. Gudjonsson, J. M. Kahlenberg, Non-lesional and lesional lupus
skin share inflammatory phenotypes that drive activation of CD16+ dendritic cells.

bioRxiv 2021.09.17.460124 [Preprint]. 20 September 2021.
https://doi.org/10.1101/2021.09.17.460124.

43. E. Der, H. Suryawanshi, P. Morozov, M. Kustagi, B. Goilav, S. Ranabathou, P. Izmirly, R. Clancy, H. M. Belmont, M. Koenigsberg, M. Mokrzycki, H. Rominieki, J. A. Graham, J. P. Rocca, N. Bornkamp, N. Jordan, E. Schulte, M. Wu, J. Pullman, K. Slowikowski, S. Raychaudhuri, J. Guthridge, J. James, J. Buyon, T. Tuschl, C. Putterman, J. Anolik, W. Apruzzese, A. Arai, C. Berthier, M. Brenner, J. Buyon, R. Clancy, S. Connery, M. Cunningham, M. Dall’Era, A. Davidson, E. Der, A. Fava, C. Fonseka, R. Furie, D. Goldman, R. Gupta, J. Guthridge, N. Hacohen, D. Hildeman, P. Hoover, R. Hsu, J. James, R. Kado, K. Kalunian, D. Kamen, M. Kretzler, H. Maecker, E. Massarotti, W. McCune, M. McMahon, M. Park, F. Payan-Schober, W. Pendergraft, M. Petri, M. Pichavant, C. Putterman, D. Rao, S. Raychaudhuri, K. Slowikowski, H. Suryawanshi, T. Tuschl, P. Utz, D. Waguespack, D. Wofsy, F. Zhang. Tubular cell and keratinocyte single-cell transcriptomics applied to lupus nephritis reveal type I IFN and fibrosis relevant pathways. Nat. Immunol. 20, 915–927 (2019).

44. T. M. Li, K. R. Veiga, N. Schwartz, Y. Chinenov, D. J. Oliver, J. Lora, A. Jabbari, Y. Liu, W. D. Shipman, M. J. Sandoval, I. F. Sollohub, W. G. Ambler, M. Rashighi, J. G. Krueger, N. Anandasabapathy, C. P. Blobel, T. T. Lu, Type I interferon modulates Langerhans cell ADAM17 to promote photosensitivity in lupus. bioRxiv 2021.08.18.456792 [Preprint]. 18 August 2021. https://doi.org/10.1101/2021.08.18.456792.

45. K. A. Kirou, C. Lee, S. George, K. Louca, M. G. E. Peterson, M. K. Crow. Activation of the interferon-α pathway identifies a subgroup of systemic lupus erythematosus patients with distinct serologic features and active disease. Arthritis Rheum. 52, 1491–1503 (2005).

46. Q. Z. Li, J. Zhou, Y. Lian, B. Zhang, V. K. Branch, F. Carr-Johnson, D. R. Karp, C. Mohan, E. K. Wakeland, N. J. Olsen. Interferon signature gene expression is correlated with autoantibody profiles in patients with incomplete lupus syndromes. Clin. Exp. Immunol. 159, 281–291 (2010).
47. E. L. Hubbard, D. S. Pisetsky, P. E. Lipsky, Anti-RNP antibodies are associated with the interferon gene signature but not decreased complement levels in SLE. *Ann. Rheum. Dis.* (2022).

48. C. C. Berthier, L. C. Tsoi, T. J. Reed, J. N. Stannard, E. M. Myers, R. Namas, X. Xing, S. Lazar, L. Lowe, M. Kretzler, J. E. Gudjonsson, J. M. Kahlenberg, Molecular profiling of cutaneous lupus lesions identifies subgroups distinct from clinical phenotypes. *J. Clin. Med.* 8, 1244 (2019).

49. S. Zampieri, M. Alaibac, L. Iaccarino, R. Rondinone, A. Ghirardello, P. Sarzi-Puttini, A. Pesarico, A. Doria, Tumour necrosis factor α is expressed in refractory skin lesions from patients with subacute cutaneous lupus erythematosus. *Ann. Rheum. Dis.* 65, 545–548 (2006).

50. L. E. Tomalin, C. B. Russell, S. Garcet, D. A. Ewald, P. Klekotka, A. Nirula, H. Norsgaard, M. Suárez-Fariñas, J. G. Krueger, Short-term transcriptional response to IL-17 receptor-A antagonism in the treatment of psoriasis. *J. Allergy Clin. Immunol.* 145, 922–932 (2020).

51. M. Robert, P. Miossec, Interleukin-17 and lupus: Enough to be a target? For which patients?. *Lupus* 29, 6–14 (2020).

52. A study to assess the safety and efficacy of secukinumab in alleviating symptoms of discoid lupus erythematosus. *U.S. Natl. Libr. Med. Clin. Trials* (2021).

53. B. Ungar, A. B. Pavel, R. Li, G. Kimmel, J. Nia, P. Hashim, H. J. Kim, M. Chima, A. S. Vekaria, Y. Estrada, H. Xu, X. Peng, G. K. Singer, D. Baum, Y. Mansouri, M. Taliercio, E. Guttman-Yassky, Phase 2 randomized, double-blind study of IL-17 targeting with secukinumab in atopic dermatitis. *J. Allergy Clin. Immunol.* 147, 394–397 (2021).

54. S. Tyring, A. Gottlieb, K. Papp, K. Gordon, C. Leonardi, A. Wang, D. Lalla, M. Woolley, A. Jahreis, R. Zitnik, D. Cell, R. Krishnan, Etanercept and clinical outcomes, fatigue, and depression in psoriasis: Double-blind placebo-controlled randomised phase III trial. *Lancet* 367, 29–35 (2006).
55. K. Reich, F. O. Nestle, K. Papp, J. P. Ortonne, R. Evans, C. Guzzo, S. Li, L. T. Dooley, C. E. M. Griffiths, Infliximab induction and maintenance therapy for moderate-to-severe psoriasis: A phase III, multicentre, double-blind trial. *Lancet* **366**, 1367–1374 (2005).

56. A. Menter, S. K. Tyring, K. Gordon, A. B. Kimball, C. L. Leonardi, R. G. Langley, B. E. Strober, M. Kaul, Y. Gu, M. Okun, K. Papp, Adalimumab therapy for moderate to severe psoriasis: A randomized, controlled phase III trial. *J. Am. Acad. Dermatol.* **58**, 106–115 (2008).

57. A. Blauvelt, K. Reich, M. Lebwohl, D. Burge, C. Arendt, L. Peterson, J. Drew, R. Rolleri, A. B. Gottlieb, Certolizumab pegol for the treatment of patients with moderate-to-severe chronic plaque psoriasis: Pooled analysis of week 16 data from three randomized controlled trials. *J. Eur. Acad. Dermatol. Venereol.* **33**, 546–552 (2019).

58. A. Lorenzo-Vizcaya, D. A. Isenberg, The use of anti-TNF-alpha therapies for patients with systemic lupus erythematosus. Where are we now?. *Expert Opin. Biol. Ther.* **21**, 639–647 (2021).

59. A. Jacobi, C. Antoni, B. Manger, G. Schuler, M. Hertl, Infliximab in the treatment of moderate to severe atopic dermatitis. *J. Am. Acad. Dermatol.* **52**, 522–526 (2005).

60. N. Cassano, F. Loconsole, C. Coviello, Infliximab in recalcitrant severe atopic eczema associated with contact allergy. *Int. J. Immunopathol. Pharmacol.* **19**, 237–240 (2006).

61. M. Yuzaiful, M. Yusof, M. Wittmann, C. Fernandez, D. Wilson, S. Edward, G. Abignano, A. Alase, P. Laws, M. Goodfield, P. Emery, E. Vita, Targeted therapy using intradermal injection of etanercept for remission induction in discoid lupus erythematosus (TARGET-DLE): Results from a proof-of-concept phase II trial. *Lupus Sci. Med.* **6**, A1–A227 (2019).

62. Targeted therapy using intradermal injection of etanercept for remission induction in discoid lupus erythematosus (TARGET-DLE). *U.S. Natl. Libr. Med. Clin. Trials* (2019).

63. A. B. Gottlieb, A. M. Goldminz, Ustekinumab for psoriasis and psoriatic arthritis. *J. Rheumatol.* **39**, 86–89 (2012).
64. A study of ustekinumab in participants with active systemic lupus erythematosus (2021); https://clinicaltrials.gov/ct2/show/NCT03517722?term=ustekinumab&cond=lupus&draw=2 &rank=3.

65. Janssen Pharmaceuticals, Janssen announces discontinuation of phase 3 LOTUS study evaluating ustekinumab in systemic lupus erythematosus (2020); www.jnj.com/janssen-announces-discontinuation-of-phase-3-lotus-study-evaluating-ustekinumab-in-systemic-lupus-erythematosus.

66. R. F. van Vollenhoven, B. H. Hahn, G. C. Tsokos, C. L. Wagner, P. Lipsky, Z. Touma, V. P. Werth, R. M. Gordon, B. Zhou, B. Hsu, M. Chevrier, M. Triebel, J. L. Jordan, S. Rose, Efficacy and safety of ustekinumab, an IL-12 and IL-23 inhibitor, in patients with active systemic lupus erythematosus: Results of a multicentre, double-blind, phase 2, randomised, controlled study. Lancet 392, 1330–1339 (2018).

67. W. D. Shipman, S. Chyou, A. Ramanathan, P. M. Izmirly, S. Sharma, T. Pannellini, D. C. Dasoveanu, X. Qing, C. M. Magro, R. D. Granstein, M. A. Lowes, E. G. Pamer, D. H. Kaplan, J. E. Salmon, B. J. Mehrara, J. W. Young, R. M. Clancy, C. P. Blobel, T. T. Lu, A protective Langerhans cell keratinocyte axis that is dysfunctional in photosensitivity. Sci. Transl. Med. 10, eaap9527 (2018).

68. S. Tian, J. G. Krueger, K. Li, A. Jabbari, C. Brodmerkel, M. A. Lowes, M. Suárez-Fariñas, Meta-analysis derived (MAD) transcriptome of psoriasis defines the “Core” pathogenesis of disease. PLOS ONE 7, e44274 (2012).

69. L. Möbus, E. Rodriguez, I. Harder, D. Stölzl, N. Boraczynski, S. Gerdes, A. Kleinheinz, S. Abraham, A. Heratizadeh, C. Handrick, E. Haufe, T. Werfel, J. Schmitt, S. Weidinger, Atopic dermatitis displays stable and dynamic skin transcriptome signatures. J. Allergy Clin. Immunol. 147, 213–223 (2021).

70. L. C. Tsoi, M. T. Patrick, S. Shuai, M. K. Sarkar, S. Chi, B. Ruffino, A. C. Billi, X. Xing, R. Uppala, C. Zang, J. Fullmer, Z. He, E. Maverakis, N. N. Mehta, B. E. Perez White, S. Getsios, Y. Helfrich, J. J. Voorhees, J. M. Kahlenberg, S. Weidinger, J. E. Gudjonsson,
Cytokine responses in nonlesional psoriatic skin as clinical predictor to anti-TNF agents. *J. Allergy Clin. Immunol.* **75**, 15–18 (2021).

71. E. Eisenberg, E. Y. Levanon, Human housekeeping genes, revisited. *Trends Genet.* **29**, 569–574 (2013).

72. S. Hänzelmann, R. Castelo, J. Guinney, GSVA: Gene set variation analysis for microarray and RNA-Seq data (2013); [www.biomedcentral.com/1471-2105/14/7](http://www.biomedcentral.com/1471-2105/14/7).

73. C. Cheadle, M. P. Vawter, W. J. Freed, K. G. Becker, Analysis of microarray data using z score transformation. *J. Mol. Diagn.* **5**, 73–81.

74. J. Menche, E. Guney, A. Sharma, P. J. Branigan, M. J. Loza, F. Baribaud, R. Dobrin, A. L. Barabási, Integrating personalized gene expression profiles into predictive disease-associated gene pools. *npj Syst. Biol. Appl.* **3**, 10 (2017).

75. M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. M. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A. K. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P. H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. Von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. Von Heijne, J. Nielsen, F. Pontén, Tissue-based map of the human proteome. *Science* **347**, 394 (2015).

76. A. Gazel, P. Ramphal, M. Rosdy, B. De Wever, C. Tornier, N. Hosein, B. Lee, M. Tomic-Canic, M. Blumenberg, Transcriptional profiling of epidermal keratinocytes: Comparison of genes expressed in skin, cultured keratinocytes, and reconstituted epidermis, using large DNA microarrays. *J. Invest. Dermatol.* **121**, 1459–1468 (2003).

77. L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, *Classification and Regression Trees* (1984).

78. L. Breiman, Random forests. *Mach. Learn.* **45**, 5–32 (2001).
79. R. Blagus, L. Lusa, SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics* **14**, 106 (2013).

80. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).