Integrated, multicohort analysis reveals unified signature of systemic lupus erythematosus

Winston A. Haynes,1,2 D. James Haddon,1,3 Vivian K. Diep,1,3 Avani Khatri,1,3 Erika Bongen,1,3 Gloria Yiu,1,3 Imelda Balboni,4 Christopher R. Bolen,1 Rong Mao,1,3 Paul J. Utz,1,3 and Purvesh Khatri1,2

1Institute for Immunity, Transplantation and Infection, 2Division of Biomedical Informatics Research, 3Division of Immunology and Rheumatology, Department of Medicine, and 4Division of Allergy, Immunology and Rheumatology, Department of Pediatrics, Stanford University School of Medicine, Stanford, California, USA.

Systemic lupus erythematosus (SLE) is a complex autoimmune disease that follows an unpredictable disease course and affects multiple organs and tissues. We performed an integrated, multicohort analysis of 7,471 transcriptomic profiles from 40 independent studies to identify robust gene expression changes associated with SLE. We identified a 93-gene signature (SLE MetaSignature) that is differentially expressed in the blood of patients with SLE compared with healthy volunteers; distinguishes SLE from other autoimmune, inflammatory, and infectious diseases; and persists across diverse tissues and cell types. The SLE MetaSignature correlated significantly with disease activity and other clinical measures of inflammation. We prospectively validated the SLE MetaSignature in an independent cohort of pediatric patients with SLE using a microfluidic quantitative PCR (qPCR) array. We found that 14 of the 93 genes in the SLE MetaSignature were independent of IFN-induced and neutrophil-related transcriptional profiles that have previously been associated with SLE. Pathway analysis revealed dysregulation associated with nucleic acid biosynthesis and immunometabolism in SLE. We further refined a neutropoiesis signature and identified underappreciated transcripts related to immune cells and oxidative stress.

In our multicohort, transcriptomic analysis has uncovered underappreciated genes and pathways associated with SLE pathogenesis, with the potential to advance clinical diagnosis, biomarker development, and targeted therapeutics for SLE.

Introduction

Systemic lupus erythematosus (SLE) is a complex, heterogeneous, chronic autoimmune disease that can affect multiple organs and tissues, including the skin, kidneys, joints, lungs, blood, and CNS. SLE follows an unpredictable disease course, punctuated by periods of flare and remission (1). High-titer, class-switched antibodies that bind to nuclear antigens, including dsDNA, ribonucleoprotein (RNP), Smith, SSA (Ro), and SSB (La), are used in the diagnosis and monitoring of SLE and are thought to be pathogenic. The heterogeneity of SLE makes it challenging for clinicians to manage. Identification of robust molecular changes associated with SLE, despite the patient heterogeneity, will likely improve our understanding and management of SLE.

A number of gene expression studies have shed light on the molecular pathogenesis of SLE. For example, microarray analyses of blood cells derived from patients with SLE have shown that the IFN pathway is dysregulated in a subset of individuals who have more active and severe disease (2–5). Increases in IFN-related genes have also been observed in subsets of patients with other diseases, including systemic sclerosis (SSc), dermatomyositis (DM), polymyositis (PM), primary Sjögren's syndrome (SS), and rheumatoid arthritis (RA), although levels of IFN-inducible gene products were typically highest in SLE (6–10). A review of the biomedical literature identified IFN and neutrophils as major focuses of recent SLE research, with approximately 150 and 40 references per year, respectively. In addition to the IFN signature, upregulation of transcripts associated with granulopoiesis and plasmablasts were observed in individuals who have SLE and were found to be associated with disease activity (3, 5). McKinney et al. used gene expression analysis of purified immune cell populations to identify a transcriptional signature in CD8+ T cells that was associated with increased likelihood of SLE disease flare (11). They went on to identify an exhaustion...
signature, associated with decreased risk of flare, in CD8+ T cells from individuals who have SLE (12). However, the majority of these studies have been limited by small sample sizes, low levels of clinical and geographic heterogeneity, potential artifacts related to use of a single experimental gene array platform, and lack of external validation. A more robust approach is needed to interrogate the molecular signatures that underlie the highly variable presentation and course of SLE.

We have previously described a multicohort analysis framework (MetaIntegrator) to identify robust disease signatures, and we have repeatedly demonstrated its applications for discovering diagnostics, prognostics and drug targets, and drug repurposing, which leverages the biological and technical heterogeneity present in the large amounts of publicly available gene expression data across a broad spectrum of conditions including infections, organ transplant, vaccination, cancer, and autoimmune diseases (13–15). MetaIntegrator is based on a random-effects meta-analysis, drawing statistical power from the integration of many diverse data sets (14). By computing effect sizes for each data set independently, MetaIntegrator embraces heterogeneity and avoids the limitations of batch effect correction. We have demonstrated application of this framework across a broad spectrum of diseases, including cancer (16, 17), solid organ transplant (13), sepsis (18), viral infection (19), tuberculosis (20), neurodegenerative diseases (21), vaccination (22), and SSc (23). Here, we applied the framework to analyze 40 publicly available whole transcriptome profile data sets containing 7471 samples from patients with SLE, individuals with other autoimmune diseases or infections, and healthy volunteers. Together, these data sets represented real-world diversity because of both (a) the biological heterogeneity, as the samples were collected from multiple tissue and cell types (e.g., blood, skin, and kidney) at 17 centers across 5 countries, and (b) the technical heterogeneity, since data were generated using diverse microarray platforms (e.g., Affymetrix arrays, Illumina beadchips, and Hitachisoft chips). Our analysis identified a robust SLE MetaSignature that (a) distinguishes SLE from other autoimmune and inflammatory diseases; (b) is present in multiple affected tissues and immune cell subsets; (c) is independent of age; and (d) is correlated with disease activity. We validated the SLE MetaSignature using additional independent publicly available transcript data sets. We then devised a custom, microfluidic quantitative PCR (qPCR) assay to analyze RNA transcripts in blood derived from a prospective, independent pediatric SLE (pSLE) cohort. Pathway analysis identified potentially novel dysregulated pathways in SLE, including those related to nucleotide biosynthesis and metabolism. Importantly, we identified a non-IFN component of the SLE MetaSignature that correlated more positively with disease activity measures than the IFN-related genes. Finally, our results discovered 14 “non-IFN, non-neutrophil” genes as underappreciated targets for biomarker and therapeutic development.

Results

Identification of the SLE MetaSignature. To perform a comprehensive, unbiased study of the molecular changes underlying SLE, we identified and downloaded gene expression data from all publicly available human SLE data sets in Gene Expression Omnibus (24). In total, we identified 40 data sets from 17 centers in 5 countries composed of 7471 samples derived from whole blood, peripheral blood mononuclear cells (PBMCs), kidney, skin, synovium, B cells, T cells, monocytes, neutrophils, and endothelial progenitor cells (Figure 1 and Tables 1, 2, 3). We randomly selected 6 data sets consisting of 370 whole blood and PBMC samples as “Discovery” data sets, based on our previous finding that 5 data sets with 250–300 samples are sufficient to find a robust disease gene signature using our multicohort analysis framework (14). We divided the remaining 34 data sets into “Validation” (2,407 samples in 8 data sets) and “Extended Validation” data sets (4,694 samples in 26 data sets). Discovery and Validation data sets were required to include PBMC or whole blood samples from healthy controls and patients with SLE. Extended Validation data sets included samples from other tissues or cell types, comparisons between SLE and other diseases, and longitudinal SLE samples.

We identified 93 significantly differentially regulated genes (82 upregulated and 11 downregulated) (Supplemental Table 1; supplemental material available online with this article; https://doi.org/10.1172/jci.insight.122312DS1) with a FDR less than or equal to 5% and an absolute effect size greater than or equal to 1 compared with healthy volunteers in the Discovery data sets (Figure 2A and Supplemental Table 1). We defined these 93 genes as the “SLE MetaSignature.” In the Validation data sets, 73 of these 93 SLE MetaSignature genes met the same filtering criteria (|ES| ≥ 1 and FDR ≤ 5%) and effect sizes for all 93 genes exhibited the same directionality as in the Discovery data sets (Figure 2B and Supplemental Figure 1). Of the 20 SLE MetaSignature genes that did not meet the filtering criteria, 18 were statistically significant (FDR ≤ 5%) but had an effect size less than 1 (median effect size, 0.78). In the Extended Validation data sets containing 7471 samples from patients with SLE, individuals with other autoimmune diseases or infections, and healthy volunteers. Together, these data sets represented real-world diversity because of both (a) the biological heterogeneity, as the samples were collected from multiple tissue and cell types (e.g., blood, skin, and kidney) at 17 centers across 5 countries, and (b) the technical heterogeneity, since data were generated using diverse microarray platforms (e.g., Affymetrix arrays, Illumina beadchips, and Hitachisoft chips). Our analysis identified a robust SLE MetaSignature that (a) distinguishes SLE from other autoimmune and inflammatory diseases; (b) is present in multiple affected tissues and immune cell subsets; (c) is independent of age; and (d) is correlated with disease activity. We validated the SLE MetaSignature using additional independent publicly available transcript data sets. We then devised a custom, microfluidic quantitative PCR (qPCR) assay to analyze RNA transcripts in blood derived from a prospective, independent pediatric SLE (pSLE) cohort. Pathway analysis identified potentially novel dysregulated pathways in SLE, including those related to nucleotide biosynthesis and metabolism. Importantly, we identified a non-IFN component of the SLE MetaSignature that correlated more positively with disease activity measures than the IFN-related genes. Finally, our results discovered 14 “non-IFN, non-neutrophil” genes as underappreciated targets for biomarker and therapeutic development.

Results

Identification of the SLE MetaSignature. To perform a comprehensive, unbiased study of the molecular changes underlying SLE, we identified and downloaded gene expression data from all publicly available human SLE data sets in Gene Expression Omnibus (24). In total, we identified 40 data sets from 17 centers in 5 countries composed of 7471 samples derived from whole blood, peripheral blood mononuclear cells (PBMCs), kidney, skin, synovium, B cells, T cells, monocytes, neutrophils, and endothelial progenitor cells (Figure 1 and Tables 1, 2, 3). We randomly selected 6 data sets consisting of 370 whole blood and PBMC samples as “Discovery” data sets, based on our previous finding that 5 data sets with 250–300 samples are sufficient to find a robust disease gene signature using our multicohort analysis framework (14). We divided the remaining 34 data sets into “Validation” (2,407 samples in 8 data sets) and “Extended Validation” data sets (4,694 samples in 26 data sets). Discovery and Validation data sets were required to include PBMC or whole blood samples from healthy controls and patients with SLE. Extended Validation data sets included samples from other tissues or cell types, comparisons between SLE and other diseases, and longitudinal SLE samples.

We identified 93 significantly differentially regulated genes (82 upregulated and 11 downregulated) (Supplemental Table 1; supplemental material available online with this article; https://doi.org/10.1172/jci.insight.122312DS1) with a FDR less than or equal to 5% and an absolute effect size greater than or equal to 1 compared with healthy volunteers in the Discovery data sets (Figure 2A and Supplemental Table 1). We defined these 93 genes as the “SLE MetaSignature.” In the Validation data sets, 73 of these 93 SLE MetaSignature genes met the same filtering criteria (|ES| ≥ 1 and FDR ≤ 5%) and effect sizes for all 93 genes exhibited the same directionality as in the Discovery data sets (Figure 2B and Supplemental Figure 1). Of the 20 SLE MetaSignature genes that did not meet the filtering criteria, 18 were statistically significant (FDR ≤ 5%) but had an effect size less than 1 (median effect size, 0.78). In the Extended Validation data
sets, which included data from diverse sample types and other diseases, the SLE MetaSignature gene effect sizes were consistent with the Discovery data set (Figure 2C). Regardless of the genetic background of the patients, technical variation, tissue, and cell type, the genes comprising the SLE MetaSignature were all differentially expressed (Figure 2, A–C), demonstrating the robustness of the SLE MetaSignature.

We defined an “SLE MetaScore” for each sample using the 93-gene signature (see Methods). In the Discovery data sets, the SLE MetaScore distinguished SLE patient samples from healthy samples with a summary area under the receiver operating characteristic curve (AUROC) of 0.95 (95% CI, 0.83–0.99) (Figure 2D). The SLE MetaScore distinguished samples from patients with SLE and healthy volunteers with high accuracy in the 8 Validation data sets (summary AUROC = 0.94; 95% CI, 0.89–0.97) (Figure 2E), further demonstrating the robustness of the SLE MetaSignature.

Of the 93 genes in the SLE MetaSignature, 46 had been previously associated with SLE (2, 3, 5, 25). To the best of our knowledge, the remaining 47 genes have not previously been associated with SLE. We performed pathway analysis of the SLE MetaSignature using Differential Expression Analysis for Pathways (DEAP) (26) to identify biological processes that are dysregulated in SLE. DEAP takes advantage of the meta-analysis effect sizes for all genes (not just those in the SLE MetaSignature) and pathway topology to identify patterns of differential expression that are consistent with known biological pathways. By taking advantage of effect sizes of all genes, DEAP significantly improves power compared with gene list–based approaches (26). Furthermore, DEAP specifies genes involved in the most differentially expressed subpathway. As input for DEAP, we used study level effect sizes from the Discovery and Validation data sets (26). Supplemental Table 2 summarizes pathways that were differentially expressed at a FDR equal to or under 10% based on 5000 random permutations of the data. In addition to the expected inflammatory pathways (e.g., IFN-γ signaling pathway, chemokine/cytokine-mediated
signaling pathway, and IL signaling pathway), our analysis identified several highly significant, unexpected pathways (salvage pyrimidine deoxyribonucleotides, formyltetrahydrofolate biosynthesis, and salvage pyrimidine ribonucleotides) related to nucleic acid metabolism. Thus, pathway analysis of the SLE MetaSignature provided insights into the biological mechanisms underlying SLE.

SLE MetaScore distinguishes SLE from other autoimmune, inflammatory, and infectious diseases. We compared SLE MetaScores across inflammatory conditions, including other autoimmune and infectious diseases, to explore its specificity to SLE. We found that adult SLE (aSLE) and patients with pSLE had significantly higher SLE MetaScores than individuals with staphylococcal infection, streptococcal pharyngitis, Still’s disease (systemic onset juvenile idiopathic arthritis; sJIA), RA, pyogenic pyoderma gangrenosum and acne (PAPA), B cell deficiency, diabetes, HIV infection, and liver transplant acute rejection in whole blood and PBMC samples across multiple independent data sets (Figure 3, A and B, and Supplemental Figure 2, A and B). In concordance with the previously reported increased severity of disease observed in patients with pSLE compared with adults (27), we found that patients with pSLE had significantly higher SLE MetaScores compared with patients with aSLE (Figure 3B). Taken together, these results demonstrate that, both in adult and pediatric populations, the SLE MetaScore is highly specific to SLE compared with other autoimmune, inflammatory, and infectious diseases.

The SLE MetaScore is systemically higher across tissues in patients with SLE. SLE is a systemic autoimmune disease that affects multiple tissues and organs. Therefore, we explored whether the SLE MetaScore is persistent in tissues other than whole blood and PBMCs in patients with SLE. SLE MetaScores were higher in a data set derived from glomeruli and tubulointerstitium of kidneys from individuals with SLE compared with pretransplant living donors (Figure 3C and Supplemental Figure 3A). SLE MetaScores were higher in a data set from synovial biopsies of patients with SLE compared with those with microcrystalline arthritis

Table 1. SLE discovery data set summaries

<table>
<thead>
<tr>
<th>Data set name</th>
<th>Tissue type</th>
<th>Center</th>
<th>Sample size</th>
<th>Discovery samples</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE1909</td>
<td>PBMC</td>
<td>Baylor University, Texas, USA</td>
<td>175</td>
<td>75</td>
<td>(123)</td>
</tr>
<tr>
<td>GSE17755</td>
<td>PBMC</td>
<td>Wakayama Medical University, Osaka, Japan</td>
<td>244</td>
<td>44</td>
<td>(124)</td>
</tr>
<tr>
<td>GSE22098</td>
<td>Whole blood</td>
<td>Baylor University, Texas, USA</td>
<td>274</td>
<td>83</td>
<td>(38)</td>
</tr>
<tr>
<td>GSE39088</td>
<td>Whole blood</td>
<td>Universite Catholique de Louvain, Brussels, Belgium</td>
<td>142</td>
<td>60</td>
<td>(125)</td>
</tr>
<tr>
<td>GSE50635</td>
<td>Whole blood</td>
<td>Mayo Clinic, Minnesota, USA</td>
<td>49</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>GSE8650</td>
<td>PBMC</td>
<td>Baylor University, TX, Texas, USA</td>
<td>246</td>
<td>59</td>
<td>(126)</td>
</tr>
<tr>
<td><strong>6 Data sets</strong></td>
<td><strong>Whole blood and PBMC</strong></td>
<td><strong>4 Centers, 3 Countries</strong></td>
<td><strong>1,130</strong></td>
<td><strong>370</strong></td>
<td></td>
</tr>
</tbody>
</table>

More extensive descriptions in Supplement S1.

Table 2. SLE validation data set summaries

<table>
<thead>
<tr>
<th>Data set name</th>
<th>Tissue type</th>
<th>Center</th>
<th>Sample size</th>
<th>Validation samples</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE12374</td>
<td>PBMC</td>
<td>Wakayama Medical University, Osaka, Japan</td>
<td>17</td>
<td>17</td>
<td>(127)</td>
</tr>
<tr>
<td>GSE24706</td>
<td>PBMC</td>
<td>University of Texas - Southwestern, Texas, USA</td>
<td>48</td>
<td>48</td>
<td>(128)</td>
</tr>
<tr>
<td>GSE49454</td>
<td>Whole blood</td>
<td>Universite Catholique de Louvain, Brussels, Belgium</td>
<td>49</td>
<td>82</td>
<td>(129)</td>
</tr>
<tr>
<td>GSE61635</td>
<td>PBMC</td>
<td>University of Miami, Florida, USA</td>
<td>129</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td>GSE65391</td>
<td>Whole blood</td>
<td>Baylor University, TX, USA, USA</td>
<td>996</td>
<td>230</td>
<td>(5)</td>
</tr>
<tr>
<td>GSE72798</td>
<td>Whole blood</td>
<td>Universite Catholique de Louvain, Brussels, Belgium</td>
<td>82</td>
<td>26</td>
<td>(130)</td>
</tr>
<tr>
<td>GSE81622</td>
<td>PBMC</td>
<td>University of Texas- Southwestern, Texas, USA</td>
<td>55</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>GSE88884</td>
<td>PBMC</td>
<td>EliLilly and Company, Indianapolis, Indiana, USA</td>
<td>1,820</td>
<td>1,820</td>
<td>(131)</td>
</tr>
<tr>
<td><strong>8 Data sets</strong></td>
<td><strong>Whole blood and PBMC</strong></td>
<td><strong>6 Centers, 3 Countries</strong></td>
<td><strong>3,196</strong></td>
<td><strong>2,407</strong></td>
<td></td>
</tr>
</tbody>
</table>

More extensive descriptions in Supplement S1.
(gout and pseudogout), osteoarthritis (OA), RA, or seronegative arthritis (Figure 3D). Finally, we found that a data set derived from skin biopsies from individuals with discoid lupus erythematosus exhibited significantly higher SLE MetaScores than healthy volunteers and individuals with psoriasis, suggesting shared pathways between systemic and cutaneous lupus (Supplemental Figure 3B). Collectively, these results provide strong evidence that the SLE MetaScore is higher in multiple affected tissues in SLE in comparison both with healthy controls and other autoimmune diseases.

The SLE MetaScore is differentially expressed in diverse immune cell types. Multiple functional changes have been described in T cells of patients with SLE, including upregulation of costimulatory molecules, hypomethylation, increased expression of key immune-related genes (28), and aberrant signaling pathway activation downstream of TCR activation (29). We found that the SLE MetaScore was significantly higher in multiple independent data sets from CD4+ T cells of patients with SLE compared with healthy volunteers (Supplemental Figure 4, A–C) and RA patients (Figure 3E).

### Table 3. SLE extended validation data set summaries

<table>
<thead>
<tr>
<th>Data set name</th>
<th>Tissue type</th>
<th>Center</th>
<th>Sample size</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE10325</td>
<td>Sorted cells: CD4+ T cells, CD19+ B cells, myeloid</td>
<td>University of Texas - Southwestern, Texas, USA</td>
<td>67</td>
<td>(132)</td>
</tr>
<tr>
<td>GSE13887</td>
<td>T cells</td>
<td>State University New York Upstate, New York, New York, USA</td>
<td>27</td>
<td>(133)</td>
</tr>
<tr>
<td>GSE24060</td>
<td>Whole blood</td>
<td>National Institute of Environmental Health Sciences, North Carolina, USA</td>
<td>80</td>
<td>(134)</td>
</tr>
<tr>
<td>GSE26949</td>
<td>Endothelial progenitor cells</td>
<td>University of Michigan, Michigan, USA</td>
<td>12</td>
<td>(135)</td>
</tr>
<tr>
<td>GSE26950</td>
<td>Endothelial progenitor cells</td>
<td>University of Michigan, Michigan, USA</td>
<td>12</td>
<td>(135)</td>
</tr>
<tr>
<td>GSE26975</td>
<td>Neutrophils</td>
<td>University of Michigan, Michigan, USA</td>
<td>29</td>
<td>(35)</td>
</tr>
<tr>
<td>GSE27427</td>
<td>Neutrophils</td>
<td>Baylor University, Texas, USA</td>
<td>47</td>
<td>(136)</td>
</tr>
<tr>
<td>GSE29536</td>
<td>Whole blood</td>
<td>Baylor University, Texas, USA</td>
<td>410</td>
<td>(137)</td>
</tr>
<tr>
<td>GSE30153</td>
<td>B cells</td>
<td>Institut de Biologie Moléculaire et Cellulaire, Strasbourg, France</td>
<td>26</td>
<td>(138)</td>
</tr>
<tr>
<td>GSE32591</td>
<td>Kidney: tubulointerstitial and glomeruli</td>
<td>University of Michigan, Michigan, USA</td>
<td>93</td>
<td>(139)</td>
</tr>
<tr>
<td>GSE36700</td>
<td>Synovial biopsy</td>
<td>Universite Catholique de Louvain, Brussels, Belgium</td>
<td>25</td>
<td>(140)</td>
</tr>
<tr>
<td>GSE36941</td>
<td>PBMC</td>
<td>Pierre and Marie Curie University, Paris, France</td>
<td>20</td>
<td>(141)</td>
</tr>
<tr>
<td>GSE37536</td>
<td>Sorted cells: monocytes and macrophages</td>
<td>Northwestern University, Illinois, USA</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>GSE37573</td>
<td>B cells: EBV transformed</td>
<td>University of Texas- Southwestern, Texas, USA</td>
<td>104</td>
<td></td>
</tr>
<tr>
<td>GSE38351</td>
<td>PBMC</td>
<td>Deutsches Rheuma-Forschungszentrum, Berlin, Germany</td>
<td>74</td>
<td>(142)</td>
</tr>
<tr>
<td>GSE4588</td>
<td>Sorted cells: CD4+ T cells, B cells</td>
<td>Universite Catholique de Louvain, Brussels, Belgium</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>GSE46920</td>
<td>Monocytes</td>
<td>Baylor University, Texas, USA</td>
<td>12</td>
<td>(143)</td>
</tr>
<tr>
<td>GSE46923</td>
<td>Monocytes</td>
<td>Baylor University, Texas, USA</td>
<td>142</td>
<td>(143)</td>
</tr>
<tr>
<td>GSE50772</td>
<td>PBMCs</td>
<td>Genentech, California, USA</td>
<td>81</td>
<td>(25)</td>
</tr>
<tr>
<td>GSE51997</td>
<td>Sorted cells: CD4+ T cells, CD16+ monocytes, CD16- monocytes</td>
<td>Deutsches Rheuma-Forschungszentrum, Berlin, Germany</td>
<td>36</td>
<td>(144)</td>
</tr>
<tr>
<td>GSE52471</td>
<td>Skin</td>
<td>Mount Sinai, New York, USA</td>
<td>38</td>
<td>(145)</td>
</tr>
<tr>
<td>GSE55447</td>
<td>Sorted cells: CD4+ T cells, CD8+ T cells, monocytes, B cells</td>
<td>Mayo Clinic, Minnesota, USA</td>
<td>208</td>
<td>(146)</td>
</tr>
<tr>
<td>GSE72747</td>
<td>Whole blood</td>
<td>Universite Catholique de Louvain, Brussels, Belgium</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>GSE78193</td>
<td>Whole blood</td>
<td>Amgen, California, USA</td>
<td>125</td>
<td>(147)</td>
</tr>
<tr>
<td>GSE88885</td>
<td>PBMC</td>
<td>Eli Lilly and Company, Indianapolis, Indiana, USA</td>
<td>908</td>
<td>(131)</td>
</tr>
<tr>
<td>GSE88886</td>
<td>PBMC</td>
<td>Eli Lilly and Company, Indianapolis, Indiana, USA</td>
<td>418</td>
<td>(131)</td>
</tr>
</tbody>
</table>

26 Data sets: Whole blood, PBMC, sorted cells, tissues

16 Centers, 4 Countries 4,056

More extensive descriptions in Supplement S1.
Similarly, the SLE MetaScore was significantly increased in a data set from CD8+ T cells of individuals with SLE, compared with healthy volunteers (Supplemental Figure 4D).

Dysregulation of B cells is a hallmark of SLE, including autoantibody production, defective negative selection, and changes in the proportions of key B cell subpopulations (30, 31). The SLE MetaScore was less robust in data sets from B cells than T cells, classifying SLE in some data sets (Figure 3F and Supplemental Figure 4E) but not others (Supplemental Figure 4, F and G). Finally, the SLE MetaScores in data sets from monocytes and neutrophils were not significantly different between patients with SLE and healthy controls (data not shown).

The SLE MetaScore is positively correlated with disease activity and inflammation. The SLE Disease Activity Index (SLEDAI) is a standardized, albeit imperfect, measure of disease severity and activity. SLEDAI is based on the presence or absence of 24 features at the time of the visit, including arthritis, rash, fever, and increases in anti-DNA autoantibodies. It is often used by clinicians to monitor disease activity in an individual SLE patient (3).
Five independent data sets that profiled PBMC or whole blood samples from patients with SLE also reported SLEDAI scores. We observed a positive correlation between SLEDAI and the SLE MetaScore across each of the 5 data sets (Figure 4, A and B, and Supplemental Figure 5, A–D). The median correlation across these studies (correlation of 0.281) was significantly elevated compared with random gene sets ($P < 0.01$). The weakest SLEDAI correlation is observed in GSE27427 (Table 3), which contains only 18 samples and is derived from neutrophils. The positive correlation of SLEDAI with SLE MetaScore in the blood is notable, since the SLE MetaSignature was identified without considering disease activity when selecting initial data sets for discovery. Furthermore, we found that the SLE MetaScore correlated highly with individual clinical measures of systemic inflammation, including erythrocyte sedimentation rate (ESR) (Supplemental Figure 5E and ref. 32), and levels of complement C3 (Figure 4C) and C4 (Supplemental Figure 5F).
Prospective validation of SLE MetaSignature in an independent pSLE cohort. We validated the SLE MetaSignature in an independent pediatric cohort by studying RNA transcripts in whole blood samples from healthy adult controls and from pediatric patients with SLE or JIA. We selected 33 genes from the SLE MetaSignature based on their significance and availability of validated probes for measuring expression using a microfluidic qPCR array (Supplemental Table 3). Thirty genes out of 33 were significantly differentially expressed in SLE samples compared with healthy adult controls and pediatric JIA patients (FDR < 5%, Supplemental Table 3). Furthermore, the SLE MetaScores based on these 33 genes in the patients with pSLE were significantly higher than healthy adult controls and pediatric JIA patients ($P = 3.7 \times 10^{-5}$ and $1.8 \times 10^{-6}$, respectively; Figure 5A); distinguished patients with pSLE with high accuracy (AUROC = 0.94); and were positively correlated with SLEDAI (Spearman’s correlation = 0.307, $P = 0.045$; Figure 5B).
A subset of the SLE MetaSignature is not robustly induced by IFN. Dysregulation of the type I IFN pathway has been repeatedly observed in subsets of patients with SLE with active disease and is thought to be a critical mediator in disease pathology. Therefore, we explored the proportion of IFN-stimulated genes in the SLE MetaSignature. We analyzed 16 transcriptome data sets composed of 190 samples derived from primary human cells treated with type I IFN to identify a robust set of type I IFN–stimulated genes (Supplemental Table 4). Of the 93 genes in the SLE MetaSignature, 70 were significantly differentially expressed (effect size > 0.8) in primary cells stimulated by type I IFN (Figure 6 and Supplemental Table 1). The remaining 23 genes in the SLE MetaSignature had low effect sizes and high FDRs within the IFN-stimulated data sets (Supplemental Figure 6), suggesting that these 23 genes were not affected in cells exposed to type I IFN.

We separated the SLE MetaSignature into “IFN” and “non-IFN” SLE MetaSignatures and computed scores as before. Both scores distinguished patients with SLE with equally high accuracy in the validation data sets (Supplemental Table 5). In 4 of 5 data sets with SLEDAI disease severity measurements, the non-IFN SLE MetaSignature had a higher correlation with SLEDAI than the IFN SLE MetaSignature (Supplemental Table 6). Collectively, our analyses identified a clinically important, non-IFN component of the SLE MetaSignature.

The role of non-IFN MetaSignature genes in neutrophils. We used immunoStates to identify cell lineages that most highly express genes that comprise the SLE MetaSignature. We found that many of the non-IFN SLE MetaSignature genes were upregulated in neutrophils (33), consistent with prior literature implicating neutrophils in SLE (3, 34–38). Low-density granulocytes exhibit enhanced type I IFN production and NETosis, a form of neutrophil cell death implicated in SLE pathogenesis (39) in which DNA neutrophil extracellular traps (NETs) are extruded from activated neutrophils (36, 39). We identified a transcript profiling data set that compared low-density granulocytes and neutrophils from patients with SLE or healthy controls. We observed that the non-IFN SLE MetaSignature was prominently found in low-density granulocytes from patients with SLE but not in neutrophils from patients with SLE or healthy controls (ref. 35 and Figure 7A). We observed a strong correlation between neutrophil abundance and SLE MetaScore in both studies where quantitative neutrophil counts were available (Supplemental Figure 10). Collectively, these results suggest that the SLE MetaSignature genes related to neutrophils are the result of an expansion of the neutrophil compartment in patients with SLE rather than an altered expression profile in SLE neutrophils.

To further explore the role of the non-IFN genes in neutrophils, we identified 4 publicly available gene expression data sets with 84 samples that explored either NETosis or neutrophil development (Supplemental Table 7). The non-IFN SLE MetaSignature was upregulated in cell lines that were stimulated to induce both Nox-dependent and Nox-independent NETosis (Figure 7B). The non-IFN SLE MetaSignature progressively increased during intermediate stages of neutropoiesis (Figure 7C and Supplemental Figure 7, A and B). Collectively, these results indicate that a significant proportion of the non-IFN SLE MetaSignature is related to transcriptional signatures of NETosis and neutropoiesis.
Identification of underappreciated, non-IFN, nonneutrophil SLE MetaSignature genes.

IFN stimulation and neutrophil involvement explained the differential expression of 79 of the 93 genes in the SLE MetaSignature (Figure 8A and Supplemental Table 1). The remaining 14 genes (termed “Underappreciated SLE MetaSignature”; Table 4) provided an opportunity to explore potentially new disease mechanisms that underlie SLE. The underappreciated SLE MetaScore correlated more positively with disease activity measurements than the IFN SLE MetaScore in every blood-derived data set (Supplemental Table 8). Interestingly, 3 members of the metallothionein family (MT1E, MT1F, and MT1HL1) were in the underappreciated SLE MetaSignature. Metallothioneins play an important role in oxidative stress responses and the clearance of heavy metals. We identified 2 data sets in which human cell lines were exposed chronically to cadmium or acutely to zinc. The underappreciated SLE MetaSignature was significantly elevated in cells exposed to heavy metals when compared with the untreated cell lines (ref. 40 and Figure 8, B and C), providing a potential link between SLE and heavy metals, or when exposed to other environmental stimuli that induce oxidative stress. A cadre of the remaining 11 genes in the underappreciated SLE MetaSignature encode molecules with interesting functions related to immune cells, while the remainder of the genes have not been linked to SLE and have yet to be well characterized in the literature.

Discussion

Previous gene expression meta-analyses in SLE have been limited to a few experiments, lacked external validation, or did not investigate the signature’s specificity to SLE (41, 42). Our method leverages biological and technical heterogeneity to identify a robust disease signature, and it has been successful in diverse diseases that
range from cancer to autoimmunity and infection (13, 16–23). We performed a multicohort gene expression analysis of more than 7,000 samples from 40 data sets representing real-world biological heterogeneity (including genetic background, age, sex, treatment, tissue, cell type, and disease duration) and technical heterogeneity (including RNA isolation, microarray platform, sample preparation, and experimental protocol) to identify a persistent SLE MetaSignature. The robustness and reproducibility of the SLE MetaSignature demonstrate its generalizability to diverse patient populations not observed in traditional, single-cohort analyses (14).

Beyond generalizability, the SLE MetaSignature was both specific to SLE and correlated with disease activity. Since the SLE MetaSignature distinguished SLE from other diseases, such as diffuse or organ specific autoimmune diseases, inflammatory arthritides, and infectious diseases, the SLE MetaSignature identified SLE-specific disease processes instead of those that are generically dysregulated in other immune-mediated diseases. SLE-DAI is the current standard for assessing severity of SLE disease activity, although it is a qualitative, subjective, and difficult-to-reproduce measure (43). Therefore, the positive correlation between the SLE MetaScore and SLEDAI suggests that the SLE MetaScore is not only capturing disease activity, but also is quantitative and objective. Therefore, it could potentially serve as a metric of disease activity in future studies or as an exploratory outcome measure in future clinical trials. Because the SLE MetaScore includes both IFN and non-IFN genes, it expands upon the current best practices of using IFN-focused gene expression to measure quantitative disease activity. Finally, to the best of our knowledge, this is the largest analysis of SLE performed to date that demonstrates that there is a transcriptional signature systemically expressed across different cell types and tissues from patients with SLE and is distinct from other autoimmune and infectious diseases. Our work has the potential to enable more precise molecular definition of SLE that is distinct from other autoimmune diseases.

The role of IFN in SLE has been important in improving the understanding of disease pathogenesis, leading to many publications defining the mechanisms of IFN in SLE (2–5) and several promising clinical trials testing anti-IFN treatments in patients with SLE (44, 45). To explore beyond this existing knowledge about the role of IFN in SLE, we specifically separated the SLE MetaSignature into genes related to IFN and genes that were independent of IFN based on a meta-analysis of 16 transcript profiling data sets from IFN-stimulated human cells. We found that the non-IFN SLE MetaSignature was equally accurate in identifying patients with SLE and is distinct from other autoimmune and infectious diseases. Our work has the potential to enable more precise molecular definition of SLE that is distinct from other autoimmune diseases.
Neutrophils also play a critical role in SLE pathogenesis. Low-density granulocytes serve as the primary source of proinflammatory NETs (3, 34–39). We found that the non-IFN SLE MetaSignature was most elevated in mature neutrophils, which contrasted with the more immature neutropoiesis signature observed in Bennett et al. (3). The non-IFN SLE MetaSignature was also elevated both in low-density granulocytes and mature neutrophils.

### Table 4. Underappreciated genes in SLE MetaSignature

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>Description</th>
<th>Associated disease</th>
<th>Prior SLE references</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCB1</td>
<td>ATP binding cassette subfamily B member 1</td>
<td>Membrane-associated protein; a member of the superfamily of ATP-binding cassette (ABC) transporters; an ATP-dependent drug efflux pump for xenobiotic compounds with broad substrate specificity.</td>
<td>Colchicine resistance; inflammatory bowel disease</td>
<td>(148–150)</td>
</tr>
<tr>
<td>CD1C</td>
<td>CD1c molecule</td>
<td>A member of the CD1 family of transmembrane glycoproteins, which mediate the presentation of primarily lipid and glycolipid antigens of self or microbial origin to T cells.</td>
<td>Mycobacterium malmoense; foramen magnum meningioma</td>
<td>(151–153)</td>
</tr>
<tr>
<td>DSC1</td>
<td>Desmocollin 1</td>
<td>A calcium-dependent glycoprotein; a member of the desmocollin subfamily of the cadherin superfamily; an adhesive protein of the intercellular desmosome junctions that is required for cell adhesion and desmosome formation.</td>
<td>Subcorneal pustular dermatosis; a subtype of IgA pemphigus</td>
<td></td>
</tr>
<tr>
<td>ELANE</td>
<td>Elastase, neutrophil expressed</td>
<td>A serine protease that hydrolyzes many proteins in addition to elastin.</td>
<td>Cyclic neutropenia; severe congenital neutropenia, autosomal dominant</td>
<td>(72)</td>
</tr>
<tr>
<td>GPR183</td>
<td>G protein–coupled receptor 183</td>
<td>A GPCR expressed in lymphocytes; upregulated upon Epstein-Barr virus infection of primary B lymphocytes.</td>
<td>Epstein-Barr virus infection</td>
<td></td>
</tr>
<tr>
<td>GRN</td>
<td>Granulin Precursor</td>
<td>A member of the secreted, glycosylated peptide family; involved in cell growth; important in normal development, wound healing, and tumorigenesis.</td>
<td>Neuronal ceroid lipofuscinosis; Grn-related frontotemporal lobar degeneration with Tdp43 inclusions</td>
<td>(154–157)</td>
</tr>
<tr>
<td>KLRB1</td>
<td>Killer Cell Lectin Like Receptor B1</td>
<td>Important for NK cells cytotoxicity.</td>
<td>Cytomegalovirus infection; inflammatory bowel disease</td>
<td>(35, 78, 79)</td>
</tr>
<tr>
<td>LHFPL2</td>
<td>LHFPL Tetraspan Subfamily Member 2</td>
<td>A member of the lipoma HMGIC fusion partner (LHFPP) gene family, which is a subset of the superfamily of tetraspan transmembrane protein encoding genes; plays a role in female and male fertility; involved in distal reproductive tract development.</td>
<td>Deafness</td>
<td></td>
</tr>
<tr>
<td>MT1E</td>
<td>Metallothionein 1E</td>
<td>A high content of cysteine residues that bind various heavy metals; transcriptionally regulated by both heavy metals and glucocorticoids.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT1F</td>
<td>Metallothionein 1F</td>
<td>A high content of cysteine residues that bind various heavy metals; transcriptionally regulated by both heavy metals and glucocorticoids.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT1HL1</td>
<td>Metallothionein 1H Like 1</td>
<td>A high content of cysteine residues that bind various heavy metals; transcriptionally regulated by both heavy metals and glucocorticoids.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAPIL3</td>
<td>Nucleosome Assembly Protein 1 Like 3</td>
<td>A member of the nucleosome assembly protein (NAP) family.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCN2</td>
<td>Transcobalamin 2</td>
<td>A member of the vitamin B12-binding protein family; binds cobalamin and mediates the transport of cobalamin into cells.</td>
<td>Transcobalamin II deficiency; transcobalamin deficiency</td>
<td></td>
</tr>
<tr>
<td>VSIG1</td>
<td>V-Set and Immunoglobulin Domain Containing 1</td>
<td>A member of the junctional adhesion molecule (JAM) family;</td>
<td>Gastric, esophageal, and ovarian cancers</td>
<td></td>
</tr>
</tbody>
</table>

The 14 genes from the underappreciated SLE MetaSignature are listed, including their gene symbol, gene name, description, and human diseases with which these genes have associated. Bolded genes are upregulated in SLE versus healthy controls; downregulated are not bolded.
in response to NETosis-inducing stimulation. Overall, our work further refines the signature of neutropoiesis in SLE and reinforces an important role for low-density granulocytes and NETosis in SLE. Although B cells have an established role in SLE, the SLE MetaScore exhibited mixed results on available sorted B cell gene expression data sets. Due to limited availability of data, we cannot conclusively evaluate whether these challenges are the result of experimental conditions or a lack of signal in B cells.

One of the most exciting discoveries in the SLE MetaSignature is the identification of 14 genes that are unrelated to type I IFN– or neutrophil-specific gene dysregulation and are genes that, by and large, have not previously been implicated in SLE pathogenesis. These newly identified genes fall into categories that include genes with interesting known biologies that are expressed in immune cells (e.g., KLRB1, GPR183 [also called EB12], CD1C, and ELANE), genes involved in inflammation and cellular stress responses (MT1E, MT1F, and MT1HL1), and individual genes related to vitamin B12 metabolism (TCN2) and epidermal cellular integrity (DSC1).

The most striking group of genes that we identified were members of the metallothionein gene family (MT1E, MT1F, and MT1HL1). Metallothioneins are intracellular, cysteine rich, metal binding proteins involved in diverse intracellular functions that include clearance of heavy metals (cadmium, zinc, and copper) and maintenance of essential ion homeostasis (46, 47). Metallothioneins normally bind zinc (48), an important element and potent antioxidant that influences redox state, enzyme activity, gene transcription, energetic metabolism, cell cycle, cell migration, invasivity, apoptosis, and proliferation (49). Both human cell line and animal studies have indicated a role for metallothioneins in protection against cadmium toxicity (50–53). Metallothioneins can be activated by a variety of stimuli, including metal ions, cytokines, and growth factors, as well as oxidative stress and radiation (49, 54). During oxidative stress, metallothioneins are upregulated to protect the cells against cytotoxicity, radiation, and DNA damage (55–57). Interestingly, metallothionein proteins are expressed at elevated levels in the kidneys of lupus nephritis patients (58). We found that transcript profiles of human cell lines exposed acutely or chronically to heavy metals resembled the underappreciated SLE MetaSignature. We hypothesize that upregulation of metallothioneins in SLE may be a protective response to elevated oxidative stress during chronic inflammatory responses and/or exposure to environmental sources of heavy metals (59–62). The importance of metallothioneins in SLE pathology is underscored by the observation that 2 additional family members (MT1A and MT2A) are induced by IFNs and were identified in the 93-gene SLE MetaSignature.

The ELANE gene encodes neutrophil elastase (NE), a serine protease implicated in host defense and tissue injury. In addition to elastin, NE also hydrolyzes proteins within azurophil granules, extracellular matrix proteins, the outer membrane protein A (OmpA) of E. coli, and the virulence factors of other bacteria (63). In contrast to the digestive serine proteases, NE has unusually high affinity for nucleic acids (64). In naive neutrophils, NE is normally stored in azurophilic granules (65, 66). Upon activation, NE translocates from azurophilic granules to the nucleus, where it partially degrades specific histones, thereby promoting chromatin decondensation and regulating the formation of NETs (67). NE-KO mice are susceptible to bacterial and fungal infections (68, 69). Mutations in ELANE can lead to cyclic and severe congenital neutropenia (70). Furthermore, the NE enzyme may also play a role in various lung, bowel, and skin inflammatory diseases (71). Dysregulation of ELANE in SLE was previously noted in a single-cohort gene expression profile (72). Although known as a neutrophil-expressed gene, we did not identify neutrophil-specific dysregulation of ELANE in our analysis of SLE data sets, and rather than being classified under the neutrophil-related SLE MetaSignature genes, ELANE was classified as an underappreciated SLE MetaSignature gene. Unexpectedly, our further analysis of cell type expression of ELANE using immunoStates (33) indicated that ELANE is most differentially expressed in hematopoietic progenitor cells and basophils (Supplemental Figure 8). This suggests that novel functions for ELANE in other cells, in addition to neutrophils, may be involved in the pathophysiology of SLE.

DSC1 encodes a calcium-dependent glycoprotein in the desmocollin subgroup of the cadherin family. The desmocollins are critical adhesive proteins of the desmosome cell-cell junction linking epithelial cells and are required for cell adhesion and desmosome formation. DSC1 is expressed in the upper epidermis of the skin (73) and has been implicated as an autoantigen for bullous skin disease (74, 75), which is also frequently manifested in patients with SLE (75). Mice lacking DSC1 exhibit epidermal fragility accompanied by defects in epidermal barrier and differentiation (76). Neonatal mice lacking desmocollin develop epidermal lesions, and older mice develop ulcerating lesions resembling chronic dermatitis. Based on the above observations, we speculate that the abnormally low levels of DSC1 in patients with SLE lead to reduced adhesion and barrier maintenance of the upper epidermis, increasing the susceptibility to develop bullae and dermatitis.
KLRB1 (also known as CD161) encodes a C-type lectin-like receptor that is composed of a disulfide-linked homodimer of approximately 40 kDa subunits and is part of the NK gene complex (NKC) (77). KLRB1 has been previously shown to be downregulated in SLE (35, 78, 79). This gene is expressed by NK cells, subsets of γδ and αβ T cells, and invariant CD1d-specific NK T cells (80–82). The KLRB1 receptor, by interacting with its ligand LLT1 (83, 84), plays an inhibitory role in NK cell–mediated cytotoxicity and IFN-γ secretion during immune responses to pathogens (83, 84). Polymorphisms in KLRB1 are associated with structural alterations of the protein and impact its regulatory functions on NK cell homeostasis and activation (86). In contrast to its inhibitory potential in NK cells, the function of KLRB1 in T cells is less clear, with reports suggesting both coactivating (81, 85, 87) and inhibitory (88, 89) effects. CD161 (KLRB1) has been used as a marker to define Th17 and Tc17 subsets of CD4+ and CD8+ T cells that secrete the proinflammatory cytokine IL-17. However, a more recent study found that CD161-expressing T cell subsets are not all committed to the Th17 axis but are much more diverse, and that expression of CD161 identifies a transcriptional and functional phenotype shared across human T lymphocytes that is independent of both T cell receptor (TCR) expression and cell lineage (87). The dysregulation of KLRB1 in SLE may be directly linked to aberrant IFN signaling pathways and immune cell subpopulations in this disease.

GPR183 (also known as EBI2) encodes the GPCR183 that binds oysterol, the most potent of which is 7α, 25-dihydroxycholesterol (7α,25-OHC) (90). GPR183 is upregulated in a Burkitt’s lymphoma cell line upon Epstein-Barr virus infection (91), an infection that is also strongly linked to SLE (91, 92). Interestingly, GPR183 is also strongly induced in UVB-irradiated skin biopsies (93) and UV light has been postulated to induce SLE photosensitivity (94) and DNA damage–driven apoptosis (95). The GPR183 protein is a negative regulator of IFN (96). In lymphoid organs, GPR183 plays a key role in mediating the migration and antibody response of multiple immune cell types, including B cells, T cells, DCs, and monocytes (97–101). GPR183-deficient mice have fewer plasma cells, reduced antibody titres (97, 98), and diminished CD4+ splenic DCs. In another study, mice lacking GPR183 or its 7α,25-OHC ligand show defects in the trafficking of group 3 innate lymphoid cells and defects in lymphoid tissue formation in the colon (102). GPR183 has been implicated in inflammatory and autoimmune diseases, including multiple sclerosis (103), inflammatory bowel disease (104), Crohn’s disease (104), type 1 diabetes, and cancer (101). In multiple sclerosis, data from the experimental autoimmune encephalomyelitis (EAE) animal model suggest that GPR183 is a critical mediator of CNS autoimmunity and regulates the migration of autoreactive T cells into inflamed organs (105). Thus, the intriguing links between GPR183 and SLE through Epstein-Barr virus, IFN, and UV light, as well as its important functions in instructing immune cell localization and antibody response, identify GPR183 and its ligand 7α,25-OHC as potential biomarkers and/or therapeutic targets for SLE.

Research in many immunological disorders, including SLE, has recently focused on the importance of immunometabolism in disease (106, 107). In SLE, particular focus has been on T cell metabolism (mitochondria, oxidative stress, mTOR, glucose, and cholesterol pathways), with additional interest in B cells (glycolysis and pyruvate), macrophages (stress response), DCs (mTOR, fatty acids), and neutrophils (NETosis, oxidation) (107). Concordant with these prior findings, our pathway analysis (26) of the SLE MetaSignature recapitulated many similar immunometabolic pathways in SLE, including pyruvate metabolism, fructose galactose metabolism, and oxidative stress response. In addition, our pathway analysis identified many non-inflammatory pathways involved in nucleic acid metabolism (including formyltetrahydrofolate biosynthesis, salvage pyrimidine deoxyribonucleotides, salvage pyrimidine ribonucleotides, and purine metabolism). Two therapies in SLE, methotrexate (108) and leflunomide (109–111), both inhibit nucleic acid metabolism (112–115), and other new molecular entities that target these pathways are entering clinical trials. Collectively, our pathway analysis results reinforce the importance of immunometabolic pathways in SLE pathogenesis.

Arguably, our approach, which leverages heterogeneity within patient populations to identify a common transcriptional signature across SLE, is ill-suited in the era of personalized medicine. A goal of personalized medicine is to cluster heterogeneous patients into homogeneous subgroups, which does not account for the individual variations that should be targeted. The underlying assumption is that the individual variation between subgroups is likely causal, which can be targeted to improve therapy and outcomes. However, it is equally likely that the disease-causing biology may be the same across all patients, and the variation observed between patients and subgroups is a result of environmental exposures. Studying a homogeneous patient population may identify a signature that explains the variation between groups but may not be causal and therapeutically relevant. Therefore, we believe that a more suitable approach would be to complement “personalized medicine” with “precision medicine” in SLE such that it first provides a precise
molecular definition of SLE, as we have done here. This could then lead to identification of multiple drug targets and corresponding therapies, increasing the number of drugs available to treat patients with SLE.

We anticipate that the full SLE MetaSignature, and particularly the underappreciated SLE MetaSignature, will be tested in blood and tissue derived from prospectively collected SLE cohorts to identify relationships between SLE flares, clinical subgroups, and responses to newly tested therapies. Another important question is whether the proteins encoded by these genes are abnormally expressed or observed in unanticipated cell populations or tissues. Our results will help guide targeted analyses of SLE blood and kidney samples using single cell technologies such as scRNA-Seq, ATAC-Seq, Cytometry Time of Flight, Multiplexed Ion Beam Imaging, and CO-Detection by IndEXing (116). Many of these methods are being used by the Accelerating Medicines Partnership RA/SLE program to characterize human SLE tissue, with a goal to identify novel pathways and disease targets (117). Ongoing studies using CRISPR screens and IHC are interrogating the role played by these genes in cultured immune cells, as well as the effect of the underappreciated SLE MetaSignature on IFN signaling, neutrophil biology, and animal models.

Our analysis has a few limitations. First, we focused on identifying a gene signature that is conserved between cohorts and across samples and that does not identify patient subgroups. Although this is beneficial for capturing features that are consistent across populations, it is ill-suited for identifying subgroups of disease. Second, because we only used publicly available data sets, our analyses were restricted to the comparisons available in the public data, including tissues, cell types, and diseases sampled. To enable even richer analysis, we encourage the research community to contribute richly annotated data sets to the public domain. In the context of SLE, particularly important annotations — when available — include: age, sex, SLEDAI with individual components specifically recorded, drugs at the time of blood draw, drug doses and start dates, organ system involvement, and disease targets (117). Many of these methods are being used by the Accelerating Medicines Partnership RA/SLE program to characterize human SLE tissue, with a goal to identify novel pathways and disease targets (117). Ongoing studies using CRISPR screens and IHC are interrogating the role played by these genes in cultured immune cells, as well as the effect of the underappreciated SLE MetaSignature on IFN signaling, neutrophil biology, and animal models.

Recent studies have been dominated by important discoveries that link type I IFN, neutrophils, and NETs to SLE. We have identified a unified SLE MetaSignature that implicates 14 underappreciated genes in SLE pathogenesis, only 4 of which were identified through a direct PubMed search of SLE (KLRB1, GRN, CD1C, and ABCB1, with 2, 5, 7, and 9 references each, respectively). Scouring of published literature reveals connections to additional genes, including ELANE, EBI2, and LHFPL2, but none of these have garnered significant attention in SLE research. Eight of the underappreciated SLE MetaSignature genes have plausible roles in SLE because they are expressed in immune cells, skin, or stress response. Perhaps even more interesting are the 6 genes (ABCB1, GRN, LHFPL2, NAPIL3, TCN2, and VSIG1) that are not linked to the immune system, plausible pathogenic mechanisms, or autoimmune diseases. Scientists often fall prey to the “streetlight effect” — looking for answers where the light is better rather than where the truth is more likely to lie (118–121). Although many of the underappreciated SLE MetaSignature genes make mechanistic sense, we should not lose sight of the 6 genes that had previously been in the shadows but are now illuminated.

Methods
Supplemental Methods are available online with this article.

Study approval. The vast majority of the data was obtained from public repositories (NCBI GEO). For the prospective validation analysis, all subjects were recruited and all samples were collected following protocols approved by the Stanford University IRB (IRB protocol 13952, 14734).

Author contributions
WAH, PJU, and PK designed experiments. DJH, VKD, and IB conducted experiments. WAH, EB, and CRB acquired data. WAH, DJH, EB, and CRB analyzed data. IB and PJU acquired samples. WAH, DJH, AK, VKD, EB, GY, IB, CRB, RM, PJU, and PK wrote the manuscript.

Acknowledgments
We would like to thank the pSLE and JIA patients and their families, the healthy volunteers who participated in this study, Human Immune Monitoring Core (HIMC) for performing qPCR assays, and members of the Khatri and Utz laboratories. WAH was funded by the National Science Foundation Graduate Fellowship under grant no. DGE-114747 and NIH NLM T15 LM 007033. IB was supported by the NIH (grant no. K08-AI-080945), the Stanford Child Health Research Institute (Child Health Research Program Pilot Grant for Early Career Investigators), and the Arthritis Foundation (Postdoctoral Fellowship).
EB was supported by the Stanford Gabilan Graduate Fellowship in Science and Engineering and the Stanford Women and Sex Differences in Medicine (WSDM) Seed Grant. GY was supported by the Stanford Medical Scientist Training Program (NIGMS grant no. 5T32GM007365-43). PK is funded by the Bill and Melinda Gates Foundation (OPPP113682); the National Institute of Allergy and Infectious Diseases (NIAID) grants 1U19AI109662, U19AI057229, and 5R01AI125197; Department of Defense contracts W81XWH-18-1-0253 and W81XWH1910235; and the Ralph & Marian Falk Medical Research Trust. PJU is the recipient of gifts from the Henry Gustav Floren Family Trust, Elizabeth Adler, and the Baxter Foundation. PJU is supported by NIH grants NIAID U19-AI110491, Stanford Autoimmunity Center of Excellence (ACE); NIAID U1 UM1AI110498-01, ACE Collaborative Project; 1 U19-AI090019; and NIAID 1 R01 AI125197-01.

Address correspondence to: Purvesh Khatri, 279 Campus Drive, Beckman Center, B235A, Stanford, California 94305, USA. Phone: 650.497.5281; Email: pkhattri@stanford.edu. Or to: Paul J. Utz, 269 Campus Drive, CCSR Building, 2215-A, Stanford, California 94305, USA. Phone: 650.724.5421; Email: pjutz@stanford.edu.


