Background Lupus nephritis affects more than one-half of patients with SLE and is the most common serious manifestation of the disease. Lupus nephritis is more common in Hispanic and African-American patients than in those of European ancestry, and class III and IV nephritis progresses to end-stage renal disease in 10%–15% of patients within 15 years of diagnosis. Identification of markers and mechanisms of lupus nephritis could provide new approaches to predict and treat disease.

Methods To identify blood cell transcriptome biomarkers that differentiate renal and non-renal disease we performed RNA sequencing on peripheral blood samples from 15 patients with lupus nephritis and 14 patients with non-nephritis manifestations of SLE (samples represented each patient during flaring and quiescent disease states) and from 3 healthy donors. To relate gene expression to activity of nephritis, 216 longitudinal samples from 30 patients with lupus nephritis covering a median time frame of 28 months were analyzed using the Illumina HT-V4 Bead array. Serum albumin levels were documented at the time of each visit.

Results Principal component analysis of RNA sequencing data clearly differentiated SLE patients with nephritis from those without nephritis, and linear models for microarray (limma) analysis identified 153 gene transcripts differentially expressed between the two patient groups (fold change >1.5; p<0.05). U1 and U3 RNA transcripts were increased in lupus nephritis samples, and the most highly expressed transcript based on fold change was TREM4L4, encoding a protein previously identified as amplifying TLR7 signaling and promoting type I interferon production. Analysis of longitudinal microarray data in relation to serum albumin identified 120 transcripts. Those most significantly correlated with lupus nephritis activity were pituitary tumor-transforming gene 1 (PTTG1), recently identified as polymorphic and associated with SLE, uridine cytosine kinase 2 (UCK2), thioredoxin (TXN), and RNASE2. Expression of PTTG1 fluctuated over time, with elevated levels preceding the time of peak renal disease activity.

Conclusions Splicosome-associated RNAs and TREML4, a TLR7-associated gene product, may represent biomarkers of lupus nephritis, and PTTG1, the product of a lupus-associated gene reported to be involved in epithelial-mesenchymal transition, may be a novel therapeutic target associated with active nephritis. These studies provide a rich data set stimulating new understanding of mechanisms contributing to lupus nephritis.

Acknowledgements This work was supported by the Lupus Research Alliance, Pfizer-Centers for Therapeutic Innovation, and the Emerald Foundation.
Results StarShipTM was used to classify 250 PBMC (50 each of CD14 monocytes, CD19 B cells, CD4 helper T cells, CD8 T cells, and CD36 NK cells). Using dynamic spherical k-means, 6 clusters were generated that closely corresponded to the known cell types (figure 1). For comparison, hierarchical clustering and one-off spherical k-means with k set to 5 were carried out. Hierarchical clustering had an ARI of 0.45, one-off spherical k-means had an ARI of 0.89, and dynamic spherical k-means had an ARI of 0.86.

Conclusions This method can effectively partition unknown cells from scRNA-Seq data sets into biologically relevant clusters without prior knowledge of the number of cell types present. The similarity between the performance of the StarShipTM algorithm and one-off k-means, which does incorporate this prior knowledge, highlights the value of this dynamic technique. A full analysis of the AMP LN data is forthcoming.

Acknowledgments Research supported by the RILITE Foundation.

Abstract BD-09 Table 1  SLICC Classification criteria identified in CLD and NMEDW

<table>
<thead>
<tr>
<th>SLICC Criteria</th>
<th>Identified in CLD (N)</th>
<th>Identified in NMEDW (N)</th>
<th>Identified in CLD and NMEDW (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>acute cutaneous</td>
<td>435</td>
<td>425</td>
<td>98%</td>
</tr>
<tr>
<td>Chronic cutaneous</td>
<td>146</td>
<td>141</td>
<td>97%</td>
</tr>
<tr>
<td>Renal</td>
<td>182</td>
<td>118</td>
<td>65%</td>
</tr>
<tr>
<td>Serositis</td>
<td>221</td>
<td>115</td>
<td>52%</td>
</tr>
<tr>
<td>Arthritis</td>
<td>472</td>
<td>161</td>
<td>34%</td>
</tr>
<tr>
<td>Neurological</td>
<td>205</td>
<td>59</td>
<td>29%</td>
</tr>
<tr>
<td>Oral ulcers</td>
<td>281</td>
<td>46</td>
<td>16%</td>
</tr>
<tr>
<td>Alopecia</td>
<td>96</td>
<td>1</td>
<td>1%</td>
</tr>
</tbody>
</table>

LABORATORY

Anti-dsDNA Ab         | 348                   | 310                     | 89%                             |
Hemolytic Anemia      | 5                     | 4                       | 80%                             |
Complement            | 500                   | 368                     | 74%                             |
Leukopenia\*          | 508                   | 369                     | 73%                             |
Lymphopenia           | 131                   | 84                      | 64%                             |
Anti-phospholipid     | 433                   | 226                     | 52%                             |
antibody              |                       |                         |                                 |
Anti-nuclear antibody | 454                   | 98                      | 22%                             |
Thrombocytopenia      | 12                    | 2                       | 17%                             |
Coombs                | 109                   | 0                       | 0%                              |

Results StarShipTM was used to classify 250 PBMC (50 each of CD14 monocytes, CD19 B cells, CD4 helper T cells, CD8 T cells, and CD36 NK cells). Using dynamic spherical k-means, 6 clusters were generated that closely corresponded to the known cell types (figure 1). For comparison, hierarchical clustering and one-off spherical k-means with k set to 5 were carried out. Hierarchical clustering had an ARI of 0.45, one-off spherical k-means had an ARI of 0.89, and dynamic spherical k-means had an ARI of 0.86.

Conclusions This method can effectively partition unknown cells from scRNA-Seq data sets into biologically relevant clusters without prior knowledge of the number of cell types present. The similarity between the performance of the StarShipTM algorithm and one-off k-means, which does incorporate this prior knowledge, highlights the value of this dynamic technique. A full analysis of the AMP LN data is forthcoming.

Acknowledgments Research supported by the RILITE Foundation.

Background SLE is difficult to diagnose given diverse manifestations that occur over time and across care sites. Electronic health records (EHR) are now used in a majority of health care settings throughout the country, and present a rich source of information about patients which can be mined for earlier diagnosis identification, to improve quality of care, or enable clinical studies. To identify SLE patients in EHR data, we developed a rules-based algorithm based on the SLICC classification criteria and compared against a gold standard SLE patient registry data set.

Methods We identified 513 patients in the Chicago Lupus Database (CLD) fulfilling 4 or more of the ACR classification criteria for SLE who also had records in the Northwestern Medicine Electronic Data Warehouse (NMEDW). ICD-9/10 codes were used to identify clinical SLICC SLE classification criteria items. Laboratory results were identified using lab test names in combination with threshold numeric values in order to determine whether patients met the SLICC lab test classification criteria requirements.

Results As shown in table 1, of 513 patients with SLE in the CLD, we detected the following SLICC classification criteria, in the NMEDW: clinical chronic cutaneous 97%; acute cutaneous 98%; renal 63%; serositis 52%; arthritis 34%; neuro 29%; ulcers 16%; alopecia 1%; and labs dsDNA 89%; hemolytic anemia 80%; complement 74%; leukopenia/lymphopenia 73%; APL 64%; ANA 52%; thrombocytopenia 22%; Coombs 17%; Sm 0%.

Of 513 patients with SLE in the CLD based on ACR criteria, 513 had at least 1 clinical criteria, 466 had at least 1 immunologic criteria, and 471 had 4 or more criteria. Using EHR data from the NMEDW, and rules for the SLICC classification criteria that were based on ICD9/10 codes and labs and required identification of at least one clinical and one immunological criteria, we categorized 450/513 (88%) patients as having definite lupus.

Conclusions Query of patient EHR data with ICD-9/10 codes and lab tests for specific SLICC classification criteria items requires refinement to improve identification of some criteria. Using the SLICC classification rule for definite SLE, we were able to identify 88% of those with definite SLE by the ACR criteria, using ICD9/10 codes and labs. Text searching of notes (by simple string matching or natural language processing) may improve identification of individual SLICC criteria (e.g. renal biopsy) and may be critical for mining physicians’ notes for criteria that are not well documented with diagnosis codes or lab results.

Acknowledgements Funding-NIH/NIAMS R21AR074081.