T cell immunophenotypes in sarcoidosis identified by cluster analysis and a transcriptomic integration

Liao SY[1,2], Li L[1,2], Barkes, B[1], Macphail,K[1], Bowler, RP[1,2], Kaminsky, N[3], Wisniewski,SR[4], Hamzeh, NY[5], Maier LA[1,2]

[1]National Jewish Health (NJH) , Department of Medicine, Denver, CO; [2]University of Colorado Anschutz Medical Campus, Department of Medicine, Aurora, CO; [3]Yale University, Department of Medicine, New Haven, CT; [4]University of Pittsburgh, Graduate School of Public Health, Pittsburgh, PA; [5] University of Iowa, Department of Medicine, Iowa City, IA

Shu-Yi Liao, MD, MPH, ScD: LIAOS@NJHealth.org The study was funded by U01HL112707, U01 HL112695, UL1TRR002535

Rationale

- Sarcoidosis is a complex systemic disease that affects between 45 and 300 per 100,000 in the United States (Erdal BS, et.al.2012).
- The disease pathogenesis is thought to be related to immune dysfunction, including imbalances between Th1, Th17.1 (IFNγ+ cells), Th17 and regulatory T cells (Treg).
- Using unsupervised cluster analysis and clinical and transcriptomic data, our goal is to identify phenotypes with distinct disease mechanisms.

Study Population

- Sarcoidosis patients were recruited at National Jewish Health (NJH) and underwent testing and phenotyping based on the Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis (GRADS) NIH multi-center study and NJH site-protocol (Moller, et.al. 2015).

Methods

- Flow cytometry was used to obtain bronchoalveolar lavage (BAL) T cell profiling.
- RNA sequencing (RNA-seq) was obtained from BAL cells based on the GRADS protocol.
- Cluster analysis was performed using the Hierarchical Clustering on Principal Components (HCPC) method on the T cell phenotypes.
- Clinical phenotypes were compared between the clusters using multiple regression models.
- Differential expression analysis, Gene Set Enrichment Analysis (GSEA), and pathway analysis using Reactome were performed on the transcriptomic data.
- A subsequent machine learning approach was performed with the goal to identify the BAL transcription classifier for the two main immunological phenotypes identified.

Results

- 57 subjects with both BAL T cell profiling and RNA-seq data were included in the final analysis.
  - 30 females (53%) and 27 males (47%)
  - Scadding Stage 0 & I: 14 (24%); II & III: 30 (53%); IV: 13 (23%)
- **Unsupervised Cluster Analysis (Figure 1)**
  - Using Th1/Th17.1 (CD4+IFN-γ+), Th17 (CD4+IL17A+), and Treg (CD4+FOXP3+) (natural Treg, IL-10-producing type 1 Treg, and TGF-β- producing Th3) percentages, four clusters were identified.
  - Two main clusters had 35 subjects (cluster 1) and 20 subjects (cluster 2),while clusters 3 and 4 had one subject each. Compared to cluster 1, subjects in cluster 2 had higher percentages of IFN-γ+ and Treg cells (p=9.7E-10 and 0.02, respectively) as shown in Figure 1.

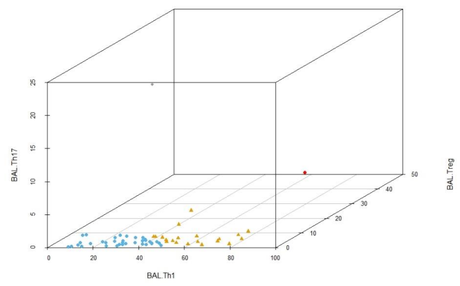

Figure 1. Clusters based on the Th profile

- **Comparing Clinical Characteristics**
  - Being in cluster 2, the odds of airway obstruction (pre-bronchodilater FEV1/FVC <0.7) increased by 6.04 (95% C.I. 1.20-40.90, p=0.04), adjusted for age, sex, race, and smoking status. With each one percent increase in Treg cells, the odds of airway obstruction increased by a factor of 1.47 (95% C.I. 1.08-2.21, P=0.03).
- **Differential Expression Analysis**
  - No differentially expressed genes were identified with a significant threshold of FDR<0.05 between cluster 1 and 2.
- **Pathway Analysis (Figure 2)**
  - Several significant pathways (FDR <0.05) were identified through GSEA and pathway analysis, including key immune system-related pathways such as IL-2 family signaling, interferon signaling and class I MHC mediated antigen-presentation.

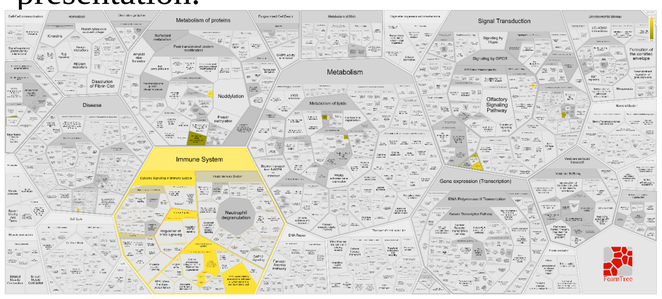

Figure 2. Pathway analysis using Reactome

- **Machine Learning with Feature Selection**
  - Poisson linear discriminant analysis (PLDA)
    - Accuracy was highest at 0.64 with a sensitivity of 0.40 and specificity of 0.77.
    - Of the 26,485 genes, 60 genes functioned as a classifier of the clusters.
      - Several of those genes are in the *HLA* region such as *HLA-DQB1*, *HLA-DQB2*.
  - Nearest shrunken centroids (NSC)
    - BAL *IGFBP2* transcription was selected as an important classifier, although this model had lower sensitivity and specificity.

Conclusion

- We identified two main immunophenotypes with distinct T cell profiling and different enrichment of gene pathways associated with interferon and IL-2 signaling, likely related to sarcoidosis pathogenesis.
- One immunophenotype (higher Tregs) was associated with a higher prevalence of airway obstruction. The immunologic~~al~~, transcriptomic and clinical features support the heterogeneity of sarcoidosis.
- Using a machine learning approach, we were able to select genes that were not previously identified through conventional differential expression analyses.

PhDPosters.com Your Research Poster, Printed For Less