Gene-Based Analysis of Sarcoidosis Susceptibility in European-Descent Americans

Liao SY[1,2,3]; Jacobson S[1]; Hamzeh, NY[4]; Culver, DA[5]; Barkes, B[1]; Mroz, P[1]; Pacheco, K[1,2,3]; Patel, D[6]; Wasfi, YS[7]; Laura Koth[8]; Langefeld, CD[9]; Maier LA[1,2,3]; Fingerlin TE[1,2,3]; and GRADs investigators[10]

[1] National Jewish Health (NJH), Department of Medicine, Denver, CO; [2] University of Colorado Anschutz Medical Campus, Department of Medicine, Aurora, CO; [3] Colorado School of Public Health, University of Colorado Denver - Anschutz Medical Campus, Aurora, CO; [4] University of Iowa, Department of Medicine, Iowa City, IA; [5] Cleveland Clinic (CC), Cleveland, OH; [6] University of Florida, Department of Medicine, Gainesville, FL; [7] Johnson & Johnson, Spring House, PA; [8] University of California-San Fransisco (UCSF), Department of Medicine, San Fransisco, CA; [9] Wake Forest School of Medicine, Department of Biostatistics and Data Science [10] Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis (GRADS)

Contact: Shu-Yi Liao, MD, MPH, ScD: LIAOS@NJHealth.org

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).
- In complex diseases like sarcoidosis, some missing heritability may be attributed to small effects of individual variants that are not generally discovered in genome-wide association studies(GWAS)
- Using a gene-based approach grouping variants within genes may help increase the ability to define genetic risk
- Furthermore, using pathway analysis may help define the potential pathogenic mechanism.

Study Population

- European-descent sarcoidosis cases and controls enrolled from National Jewish Health, Cleveland Clinic, University of California-San Francisco, and cases from Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis (GRADS) NIH multi-center study (Moller, et.al. 2015)
- 2,599 European-descent subjects, with1,335 cases (51% female) and 1264 controls (35% female)

Methods

- A GWAS was conducted using the Illumina HumanOmini 2.5 BeadChip
- Quality control measures to exclude subjects or single nucleotide polymorphisms (SNPs) included sex mismatch, genotype rate, and large departures from Hardy–Weinberg Equilibrium in controls.
- Imputed variants were tested for association with sarcoidosis under an additive model on the log-odds scale
- Gene-based analyses were performed using Versatile gene-based association study-2 (VEGAS2) using individual SNP p-values
  - Step 1: assigns SNPs to genes based on the hg19 genomic location
  - Step 2: uses a simulation approach to calculate gene-based empirical association p-values (accounting for linkage disequilibrium between SNPs)
- Downstream pathway analysis was performed using VEGAS2 with gene-sets from multiple data source
  - Gene ontology (GO)
  - MSigDB (containing canonical pathways and BIOCARTA, REACTOME, KEGG databases)
  - PANTHER
  - Pathway Commons databases
- Statistical significance thresholds were set using Bonferroni corrections
  - Genes: $p<2x10^{-6}$ (0.05/23840)
  - Pathways $p<8.4x10^{-6}$ (0.05/5959)

Results

- No single gene or pathway reached the defined significant thresholds
- Top-ranked genes ($p<10^{-4}$):
  - *PRSS1, KPNB1, LOC101927136, HLA-DRB5, PRSS3P2, ICAM5, HLA-DRA*
- Top five downstream pathways ($p<10^{-4}$):
  - TCR signaling in naïve CD4+ T cells
  - B lymphocyte cell surface molecules
  - CXCR4-mediated signaling events
  - Regulation of immune effector process
  - Perikaryon pathway

Table 1. Top 10 genes associated with sarcoidosis

Chr	Gene	Number of SNPs	Gene-based p-value	Top SNP	Top SNP p-value
7	PRSS1	11	1.10E-05	rs10246334	6.22E-05
17	KPNB1	6	2.10E-05	rs114188000	0.00067
6	LOC101927136	27	7.40E-05	rs150312573	4.51E-06
6	HLA-DRB5	48	7.50E-05	rs72851017	2.66E-05
7	PRSS3P2	10	7.50E-05	rs3857776	7.64E-05
19	ICAM5	23	8.00E-05	rs62130686	0.00013
6	HLA-DRA	5	8.80E-05	rs147430646	6.81E-06
6	HLA-DRB6	58	0.000126	rs28579642	2.28E-06
19	ICAM4	20	0.000127	rs62130686	0.00013
2	MGAT5	89	0.000139	rs1439112	0.000377

Table 2. Top 10 pathways associated with Sarcoidosis

Pathway name	P value
PID_TCR_PATHWAY	1.20E-05
BIOCARTA_BLYMPHOCYTE_PATHWAY	5.23E-05
PC_CXCR4-mediated_signaling_events	2.46E-05
GO:0043204_perikaryon	7.59E-05
GO:0002697_regulation_of_immune_effector_process	4.39E-05
REACTOME_TRANSPORT_OF_RIBONUCLEOPROTEINS_INTO_THE_HOST_NUCLEUS	0.000101904
GO:0042177_negative_regulation_of_protein_catabolic_process	0.000120432
GO:0051092_positive_regulation_of_NF-kappaB_transcription_factor_activity	0.000185675
PC_Interactions_of_Rev_with_host_cellular_proteins	0.000239807
GO:0031330_negative_regulation_of_cellular_catabolic_process	0.000241582

Conclusion

- We failed to find any statistically significant genes/pathways using this approach
- This may be due to the use of a more conservative statistical threshold and/or combining SNPs of varying association in genes
- We identified several immunologic pathways associated with sarcoidosis susceptibility, supporting that a gene-based analysis can identify candidates for future investigation
- Several novel genes outside of the *HLA* region were identified as potential candidates through the gene-based analysis, such as *PRSS*, *KPNBA*, *PRSS3P2*, and *ICAM5*. Among them, *ICAM5* has been associated with the risk and severity of idiopathic pulmonary fibrosis.