Sarcoidosis Blood Transcriptome Reflects Lung Inflammation and Overlaps with Tuberculosis

Laura L. Koth¹, Owen D. Solberg², Jeffrey C. Peng², Nirav R. Bhakta¹, Christine P. Nguyen¹, and Prescott G. Woodruff^{1,2}

¹Division of Pulmonary and Critical Care Medicine, Department of Medicine and ²Cardiovascular Research Institute, University of California, San Francisco, San Francisco, California

Rationale: Sarcoidosis is a granulomatous disease of unknown etiology, although *M. tuberculosis* may play a role in the pathogenesis. The traditional view holds that inflammation in sarcoidosis is compartmentalized to involved organs.

Objectives: To determine whether whole blood gene expression signatures reflect inflammatory pathways in the lung in sarcoidosis and whether these signatures overlap with tuberculosis.

Methods: We analyzed transcriptomic data from blood and lung biopsies in sarcoidosis and compared these profiles with blood transcriptomic data from tuberculosis and other diseases.

Measurements and Main Results: Applying machine learning algorithms to blood gene expression data, we built a classifier that distinguished sarcoidosis from health in derivation and validation cohorts (92% sensitivity, 92% specificity). The most discriminative genes were confirmed by quantitative PCR and correlated with disease severity. Transcript profiles significantly induced in blood overlapped with those in lung biopsies and identified shared dominant inflammatory pathways (e.g., Type-I/II interferons). Sarcoidosis and tuberculosis shared more overlap in blood gene expression compared with other diseases using the 86-gene signature reported to be specific for tuberculosis and the sarcoidosis signature presented herein, although reapplication of machine learning algorithms could identify genes specific for sarcoidosis.

Conclusions: These data indicate that blood transcriptome analysis provides a noninvasive method for identifying inflammatory pathways in sarcoidosis, that these pathways may be leveraged to complement more invasive procedures for diagnosis or assessment of disease severity, and that sarcoidosis and tuberculosis share overlap in gene regulation of specific inflammatory pathways.

Keywords: gene expression profiling; interferons; algorithms; computational biology

Sarcoidosis is a systemic inflammatory granulomatous disease. Regarding the immunology of sarcoidosis, studies have shown discordance between immune activation in the circulation compared with that in organs manifesting granulomatous inflammation. Specifically, many patients have anergy to skin tests (1–4) and impairment of proliferative responses by blood mononuclear cells *in vitro* (4–7). Furthermore, studies have described profound T-cell activation in the lung compared with T cells from the blood (4, 8–21), leading to the conclusion that

AT A GLANCE COMMENTARY

Current Scientific Knowledge on the Subject

Sarcoidosis is a granulomatous disease of unknown etiology, although *M. tuberculosis* may play a role in the pathogenesis. Traditional view holds that inflammation in sarcoidosis is compartmentalized to involved organs.

What This Study Adds to the Field

In this study, we provide genomic evidence that challenges current thinking about the compartmentalization of inflammation in sarcoidosis and demonstrate the significant overlap in transcriptional profiles between sarcoidosis and tuberculosis, which may have important implications for sarcoidosis etiology.

activation of T cells occurs locally (i.e., in the diseased organs). The concept that inflammatory responses are “compartmentalized” in sarcoidosis has led to the persistent belief that studies of circulating immune cells do not reflect the types of inflammatory pathways activated in diseased organs.

Two major exposures have been considered to play roles in the pathogenesis of sarcoidosis: microbial organisms and noninfectious environmental agents (inorganic or organic) (22–24). Due to the histologic similarities between tuberculosis and sarcoidosis, there have been exhaustive attempts to isolate mycobacterial organisms in tissue samples from patients with sarcoidosis, yet intact mycobacteria have not been identified conclusively. More recent work using PCR- or immune assay-based techniques have increasingly found evidence for a link between *M. tuberculosis* (TB) and sarcoidosis by identifying mycobacterial nucleic acids, insoluble mycobacterial proteins, and specific T-cell immune responses to mycobacterial peptides in many subjects with sarcoidosis (23, 25–28). Taken together, these data suggest a role for measurable immune responses to components of mycobacterium, rather than an infection *per se*, in a certain percentage of individuals who develop sarcoidosis.

A recent genomic study of subjects with active pulmonary TB (PTB) demonstrated a robust peripheral blood transcriptional signature compared with control subjects and those with latent TB (LTB) (29). This study further showed that the TB signature tracked with extent of disease on chest radiograph and reverted to that of control subjects after antimicrobial treatment. Given the evidence suggesting that mycobacterial product(s) may drive abnormal immune responses in sarcoidosis and the observation that PTB has a robust peripheral blood signature, we proposed two hypotheses: (1) that the transcriptional signature of blood would be significantly robust to distinguish patients with sarcoidosis from control subjects and identify inflammatory pathways that reflect granulomatous inflammation in involved

(Received in original form June 28, 2011; accepted in final form August 5, 2011)

Supported by grants NAIAD A1079340 and NHLBI HL09537.

Author Contributions: Conception and design, L.L.K., P.G.W., O.D.S., N.R.B.; analysis and interpretation, L.L.K., P.G.W., O.D.S., N.R.B., C.P.N., J.C.P.; drafting the manuscript for important intellectual content, L.L.K., P.G.W., O.D.S., N.R.B.

Correspondence and requests for reprints should be addressed to Laura L. Koth, 505 Parnassus Ave, Box 0111, San Francisco, CA 94143. E-mail: Laura.Koth@ucsf.edu

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Copyright ©2011 by the American Thoracic Society

Am J Respir Crit Care Med Vol 184, pp 1153–1163, 2011

Originally Published in Press as DOI: 10.1164/rccm.201106-1143OC on August 18, 2011

Internet address: www.atsjournals.org

tissues and (2) that there would be significant overlap in the blood transcriptional signature from subjects with sarcoidosis and active PTB. Our results challenge the traditional view that holds that sarcoidosis is a “compartmentalized” disease of organs by showing that transcript profiles in the blood can recapitulate inflammatory pathways in the diseased lung. In addition, data presented herein justify future studies to identify blood biomarkers of disease progression and phenotype in sarcoidosis. Some of the results of these studies have been previously reported in the form of an abstract (30).

METHODS

University of California, San Francisco Subjects: Derivation Data Set

The blood gene expression derivation set included 38 patients with sarcoidosis and 20 healthy control subjects enrolled prospectively at University of California, San Francisco (UCSF) (Table 1). The diagnosis of sarcoidosis was based upon established criteria (31). Subjects with other concurrent systemic inflammatory conditions were excluded. Control subjects without lung disease or other significant medical conditions were recruited from the community through advertisements. The study was approved by the UCSF Committee on Human Research.

Oregon Health Sciences University Subjects: Validation Data Set

The blood gene expression validation set comprised publicly available data (GEO - GSE18781, from Oregon Health Sciences University [32]) (see Table E1 in the online supplement). Review of the publication associated with this dataset indicates that all patients with sarcoidosis had symmetric hilar adenopathy as judged by chest radiograph or by CT scan. Six patients also had biopsy-proven granulomas. An additional five subjects were diagnosed based on coexisting uveitis on dilated examination because the combination of symmetric hilar

adenopathy and uveitis has been considered to be specific for a diagnosis of sarcoidosis (33). One subject was diagnosed on the basis of symmetric hilar adenopathy without uveitis or tissue confirmation (sample ID GSM465961), and analyses were performed with and without this subject in the validation set to assess the impact of this subject. No subjects were taking immunosuppressive medications. Control subjects were recruited from an ophthalmology clinic for routine eye care and had no evidence of uveitis.

TB Cohorts

Blood gene expression datasets from tuberculosis (TB) cohorts consisted of subjects with active PTB or LTB before treatment with antimicrobials and were obtained from publicly available data published by Berry and colleagues (29) (Figure 1). The groups consisted of 34 patients with PTB, 38 patients with LTB, and 24 control subjects from two London, United Kingdom cohorts (GEO accession numbers GSE19439 and GSE19444). These subjects were over 18 years of age and HIV negative, and all PTB subjects had laboratory isolation of TB on mycobacterial culture of a respiratory specimen (sputum or bronchoalveolar lavage fluid). LTB subjects were defined as a positive tuberculin-skin test or a positive TB antigen-specific IFN- γ release assay. Control subjects were negative for both assays.

Additional Cohorts Representing Infectious Disease and Th1-Type Inflammation

Additional blood gene expression datasets from disease cohorts analyzed in this study included six subjects with hypersensitivity pneumonitis from UCSF (Figure 1 and Table E2) as well as data from cohorts that were publicly available and published by Chaussabel and colleagues (34). This dataset (GEO accession number GSE22098) included 28 adult subjects with systemic lupus erythematosus (with 17 matched control subjects), 49 pediatric SLE subjects (with 19 matched control subjects), 23 subjects with group A *Streptococcus* infection, and 40 subjects with *Staphylococcus aureus* infection (together sharing a pool of 12 matched control subjects) (Figure 1).

TABLE 1. CLINICAL CHARACTERISTICS OF THE UNIVERSITY OF CALIFORNIA, SAN FRANCISCO SARCOIDOSIS COHORT

	Healthy Control Subjects	Patients with Sarcoidosis	Control vs. Sarcoidosis (<i>P</i> value)	Sarcoidosis (low lung function*)	Sarcoidosis (normal lung function)	Low vs. Normal Lung Function (<i>P</i> value)
n	20	38		18	20	
Age, yr ($\pm$ SD)	51 $\pm$ 12	50.8 $\pm$ 11.2	0.94 [†]	50.3 $\pm$ 12.2	51.2 $\pm$ 10.7	0.81 [†]
Sex (M:F)	5:15 (25% male)	17:21 (45% male)	0.17 [‡]	11:7 (61% male)	6:14 (30% male)	0.10 [‡]
Ethnicity						
Asian	0	2	0.58 [‡]	1	1	1.00 [‡]
Black	2	4		2	2	
Latino	2	1		0	1	
White	16	31		15	16	
Sarcoidosis stage (0:1:2:3:4)	NA	1:7:13:7:10		0:0:8:3:7	1:7:5:4:3	0.02 [†]
FEV ₁ , L	NA	2.56 $\pm$ 0.8		2.15 $\pm$ 0.7	2.93 $\pm$ 0.6	0.001 [†]
FEV ₁ , % predicted	NA	82.2 $\pm$ 23.3		61.6 $\pm$ 13.0	100.7 $\pm$ 12.2	2.8e-11 [†]
FVC, L	NA	3.60 $\pm$ 0.80		3.42 $\pm$ 0.90	3.75 $\pm$ 0.70	0.22 [†]
FVC, % predicted	NA	90.3 $\pm$ 16.7		77.3 $\pm$ 9.1	102.0 $\pm$ 12.6	4.7e-8 [†]
TLC, L	NA	5.47 $\pm$ 1.0		5.31 $\pm$ 1.1	5.61 $\pm$ 1.0	0.38 [†]
TLC, % predicted	NA	94.3 $\pm$ 14.9		84.6 $\pm$ 9	103.0 $\pm$ 13.7	2.1e-5 [†]
D _{LCO}	NA	21.1 $\pm$ 6.0		19.7 $\pm$ 6.2	22.5 $\pm$ 5.5	0.15 [†]
D _{LCO} , % predicted	NA	75.1 $\pm$ 20.6		63.7 $\pm$ 17.2	85.3 $\pm$ 18.1	5.9e-4
Current smoker (yes: no)	0:20	2:36	0.54 [†]	2:16	0:20	0.22 [‡]
Prednisone	NA	12 (32%)		8 (40%)	4 (22%)	0.005 [‡]
Inhaled corticosteroids	NA	9 (24%)		7 (35%)	2 (11%)	0.01 [‡]
Azathioprine	NA	1 (3%)		0	1 (6%)	0.27 [‡]
Plaquenil	NA	1 (3%)		1 (5%)	0	0.58 [‡]
Methotrexate	NA	2 (5%)		2 (10%)	0	0.47 [‡]
Mycophenolate	NA	1 (3%)		1 (5%)	0	0.58 [‡]
TNF- α blockade	NA	1 (3%)		1 (5%)	0	0.58 [‡]

Definition of abbreviations: NA = not applicable; TLC = total lung volume.

*Low lung function is defined as FEV₁ or FVC < 80% predicted. Pulmonary function measures are listed as noted $\pm$ SD.

[†] Welch two-sample *t* test.

[‡] Fisher's exact test.

Figure 1. Flow diagram of bio-informatic analyses presented in the study. (A) Overview of random forest classifier development using derivation and validation blood gene expression datasets. (B) Overview of the analyses used to compare gene expression in blood versus lung biopsies. (C) Overview of blood gene expression datasets used to qualitatively compare gene expression between cohorts of sarcoidosis, tuberculosis, hypersensitivity pneumonitis, and other infectious and Th1-type inflammatory diseases.

Lung Biopsy Cohort

We analyzed lung biopsy gene expression datasets from six patients with active pulmonary sarcoidosis and six control subjects using publically available data from Ohio (35) (GEO accession number GSE16538) (Figure 1). Clinical details reported on these subjects were that lung samples from the subjects with sarcoidosis contained well formed, non-necrotizing epithelioid granulomas in the absence of identifiable infection or foreign bodies. Control subjects had normal lung histology (35). We also analyzed lung biopsy gene expression datasets from eight subjects with “nodular self-limiting” sarcoidosis and seven subjects with “progressive-fibrotic” sarcoidosis (36). These subjects had “active” disease at the time of biopsy and were not on inhaled or oral immunosuppression, and subjects in the progressive-fibrotic group had abnormal spirometry or DL_{CO} (36). This study did not include healthy control subjects.

Gene Expression Analyses

For the blood gene expression derivation set, we isolated whole blood RNA using PAXgene RNA tubes (Becton Dickinson, Franklin Lakes, NJ) and isolation kits (PAXgene Blood RNA Kit; Qiagen, Valencia, CA). After RNA quality assessment (Agilent Bioanalyzer 2100; Agilent Technologies, Santa Clara, CA), hybridizations were performed at the

Gladstone Institute Microarray Core Facility using Affymetrix U133 Plus 2.0 microarrays (Affymetrix Inc., Santa Clara, CA). Methods for background adjustment, normalization and probe summarization, and data processing are provided in the online supplement.

Development of the Classification Algorithm

To develop a classification algorithm (patients with sarcoidosis versus control subjects) based on blood gene expression, we initially applied three different machine learning algorithms to the derivation set—random forest (37), elastic net (38), and shrunken centroids (39)—and performed subsequent analyses with random forest (R/Bioconductor) (40) because it performed best on cross-validation and provided a ranking of the genes important in classification. The Oregon blood gene expression dataset provided external validation for receiver operator characteristic (ROC) curve analyses. Additional details are provided in the online supplement.

Comparison of Blood Transcriptome with Lung Tissue

To understand how blood transcript profiles relate to those found in granulomatous tissues, we analyzed the overlap in differentially expressed genes in blood (UCSF) and lung tissue (Ohio, GSE16538) (35) from patients with sarcoidosis as compared with healthy control

subjects (Figure 1). In these analyses, we identified genes concordantly up- and down-regulated in blood and lung. To understand whether any genes in our peripheral blood transcript profile relate to fibrotic lung disease in sarcoidosis, we compared our differentially expressed genes with previously published lung biopsy data from Lockstone and colleagues (36) (Figure 1). That study identified genes that were up- or down-regulated in lung tissue from subjects with “progressive-fibrotic” as compared with “nodular self-limiting” sarcoidosis (no healthy control group was used). Overlap between genes differentially expressed in the UCSF blood dataset (patients with sarcoidosis compared with control subjects, false discovery rate (FDR) < 0.05 using limma) and genes reported as differentially expressed in lung from the “progressive-fibrotic” group in their original paper (progressive-fibrotic sarcoidosis compared with nodular self-limiting [36]) were identified.

Cross-Platform Comparisons Using Blood Gene Expression Datasets

We assembled a meta-dataset consisting of the UCSF subjects (GSE19314), the Oregon dataset (GSE18781), and the other disease control subjects described above (29, 34) using normalization approaches described in detail in the online supplement. To examine overlap between groups, we constructed heat maps and principal component analysis (PCA) plots using two prespecified gene sets: (1) the 50 most discriminative genes in the sarcoidosis classifier and (2) the 86-gene “TB-signature” identified by Berry and colleagues (29). To identify genes potentially useful in distinguishing sarcoidosis from PTB, we used a dataset that merged these two groups, trained a second random forest classifier (details are provided in the online supplement), and constructed a heat map and PCA plot using the 50 most discriminative genes. A flow diagram outlining the origin of all datasets and a summary of the bioinformatic analyses performed is provided in Figure 1.

Real-Time PCR Confirmation of Selected Genes

Two-step, real-time quantitative PCR was performed as described previously (41, 42) using primers and probes presented in Table E3.

Statistical Analysis

Comparisons of age, sex, and clinical measures were performed using the Welch two-sample *t* test and Fisher’s exact test as indicated. Classification was performed using random forests, and ROC curves were developed as described above. These analyses and PCA were performed using R/Bioconductor (43). Pathway analyses were performed using Ingenuity Pathway Analysis (<http://www.ingenuity.com>) using Canonical Pathways. Comparisons for the expression patterns of the top 10 discriminative genes were made using a pairwise *t* test with Benjamini and Hochberg’s false-discovery rate *q* values to correct for multiple comparisons (44) ($P < 0.05$ taken as statistically significant).

RESULTS

Clinical Characteristics of Subjects with Sarcoidosis

The UCSF cohort consisted of 38 prospectively enrolled patients with sarcoidosis and 20 unaffected control subjects (Table 1). The Oregon Health Sciences University sarcoidosis cohort used for validation in this study included 12 patients with sarcoidosis and 12 control subjects (GSE18781; Table E1) (referred to as “Oregon” in the figures). There were no significant differences in age or sex between the UCSF and Oregon datasets ($P > 0.05$).

Discrimination of Patients with Sarcoidosis from Control Subjects Using Machine Learning Algorithms

Sarcoidosis is a systemic disease that is widely believed to have a compartmentalized inflammatory process. However, despite this belief, two prior studies have demonstrated “activation” of peripheral blood T cells in patients with sarcoidosis compared with control subjects in the degree of effector function (45) and with regard to IL-2 receptor expression (46). Therefore, we hypothesized that a genome-wide approach to analyzing the transcriptome of whole blood cells from patients with sarcoidosis would reveal robust changes in gene expression that could distinguish patients with sarcoidosis from control subjects with high accuracy.

To do this, we took advantage of machine learning algorithms, which are particularly well suited for classification problems in which the set of features (genes) is large. We began these analyses using three independent machine learning algorithms: random forests (37), shrunken centroids (39), and elastic net (38). Each algorithm demonstrated a high sensitivity and specificity to discriminate patients with sarcoidosis from control subjects based on blood gene expression profiles, and we chose to perform all additional analyses using the random forest algorithm as described in MATERIALS AND METHODS.

We trained the random forest classifier using the UCSF microarray data (38 patients with sarcoidosis and 20 control subjects). We then assessed the classifier performance in two ways. First, prediction of class for “out-of-bag” UCSF samples (those not used for classifier development in any given classification tree) was used to provide an internal estimate of classifier performance, which was globally assessed through generation of ROC curves (Figure 2A). Then, the Oregon dataset (12 patients with sarcoidosis and 12 control subjects) was used for external validation of the classifier (2B). In the training dataset (UCSF), the error rate was 12.1% (sensitivity, 85%; specificity, 90%), estimated from the out-of-bag samples (Figure 2A). In the external validation dataset (Oregon), the observed error rate was 8% (sensitivity, 92%; specificity, 92%)

Figure 2. Receiver operator characteristic curve analysis of subject classification derived using the random forest algorithm on blood gene expression data. (A) University of California, San Francisco (UCSF) Derivation (training) set performance for all patients with sarcoidosis ($n = 38$) versus control subjects ($n = 20$). (B) Performance of the model using an external validation test set (Oregon, 12 patients with sarcoidosis and 12 control subjects). Derivation set performance is estimated using the out-of-bag error rate that is returned by the random forest algorithm, whereas the validation set performance is a comparison of the predicted versus the actual class for each subject.

(Figure 2B). All performance metrics are reported at the optimal decision threshold (i.e., the upper-left most point in the ROC curve). Analyses performed after omitting the subject in the Oregon dataset who was diagnosed with sarcoidosis on the basis of hilar adenopathy alone yielded similar results (error rate, 9%; sensitivity, 92%; specificity, 90%). These data indicate that sarcoidosis induces robust changes in the transcriptional profile of blood as compared with control subjects.

Because 12 of the 38 subjects within the UCSF sarcoidosis cohort were taking systemic corticosteroids, we assessed whether the use of corticosteroids had a significant influence on the classifier. Two observations suggest that this is not the case. First, none of these 12 subjects was misclassified on out-of-bag prediction. Second, we performed PCA using the 100 genes most important in the classification algorithm based on their value in discriminating groups (defined in random forests as the “Gini coefficient” for each gene). Plotting the first and second principal components showed that patients with sarcoidosis who were using corticosteroids were intermixed with patients who were not and that both groups were well distinguished from healthy control subjects (Figure 3A), indicating that the discriminative power of the genes used in the classifier is not specifically influenced by corticosteroid use. To understand if a smaller number of genes could retain the predictive value, we repeated the PCA using the 10 most important genes in our random forest algorithm (Figure 3B). Again, the patients with sarcoidosis remained well distinguished from healthy control subjects. Similar retention of predictive value was observed in the Oregon dataset after reduction in the number of genes used for classification (Figures 3C and 3D).

Validation of the Most Discriminative Genes

Quantitative PCR was performed for the 10 most discriminative genes (Table 2) using samples from the UCSF dataset to validate microarray findings, and in each instance PCR results correlated extremely well with microarray data (Figure 4).

Relationship with Sarcoidosis Lung Disease Severity

To determine whether the most highly discriminative genes varied with disease severity, we analyzed gene expression levels for the

top 10 discriminative genes in subjects with sarcoidosis and low lung function (FEV_1 and/or $FVC < 80\%$ predicted) separately from those with normal lung function in the UCSF dataset (Figure 5). Among the 10 genes, 3 are IFN inducible or encode for mediators downstream of the IFN- γ signaling pathway. One of these genes, *IRF1*, was statistically significantly different between patients with sarcoidosis with low and normal lung function. Although we do not have lung function data for the Oregon dataset, these analyses demonstrate the consistency of gene expression across the UCSF and Oregon datasets (Figure 5).

Similarities of Pathway Analyses in Blood and Granulomatous Lung Tissue

To understand how peripheral blood transcript profiles relate to those found in granulomatous tissues, we analyzed the overlap in differentially expressed genes in blood and lung tissue (GSE16538) (35) from patients with sarcoidosis (Figure 6). These data demonstrate that genes differentially expressed in both compartments are more likely than chance to be concordantly up-regulated and, in particular, that those with high fold-change are concordantly up-regulated in blood and lung. Specifically, we defined concordantly up-regulated genes as those with an FDR q value of < 0.1 in both datasets and fold-change greater than 2 in either dataset (a relatively lenient FDR threshold was chosen to optimize power in the biopsy analysis, which had a small sample size, but this threshold was applied across both datasets and to up- and down-regulated genes to avoid bias in the comparison). The results showed five genes significantly down-regulated in both datasets, two genes up-regulated in the whole blood dataset but down-regulated in the biopsy dataset, 10 genes down-regulated in the whole blood dataset but up-regulated in the biopsy dataset, and 33 genes up-regulated in both datasets. This excess in concordantly up-regulated genes in both datasets was statistically significant (Fisher's exact test; $P = 3.3e-10$). Among the genes that were concordantly induced in the lung and blood in sarcoidosis was a critical transcriptional regulator in the IFN- α , - β , and - γ signaling pathways (*STAT1*). Further, up-regulation of genes related to the IFN signaling pathways was significantly overrepresented in the blood transcriptome from patients with

Figure 3. Principal component analysis (PCA) using (A) the 100 genes from blood most important in the classification algorithm based on their value in discriminating groups (defined in random forests as the Gini coefficient). Patients with sarcoidosis who were taking corticosteroids were intermixed with patients who were not, and both groups were well distinguished from healthy control subjects. PC1 and PC2 denote the first and second principal components. (B) PCA using the 10 most important genes in the classification algorithm suggesting that smaller subsets of genes have the potential to retain discriminative capacity. (C and D) Similar findings were derived using PCA of the Oregon dataset.

Figure 4. Quantitative PCR confirmation of array-based blood gene expression measurement from 10 of the most important classifier genes (as ranked by the Gini coefficient). The correlation of array to quantitative PCR data was statistically significant for all genes.

sarcoidosis (Figure E1 and Table E4). Collectively, these data are consistent with a large body of literature showing a significant role for Th1-type inflammation in sarcoidosis (reviewed in Ref. 47) and demonstrate that gene regulation of specific inflammatory pathways is overrepresented in circulating blood cells and

may serve as a surrogate marker for monitoring disease severity (Figure 5) and activity.

In addition to IFN pathway-related transcripts in blood, we found up-regulation of important genes related to T-cell homeostasis and survival (*IL15* and *IL7R* [CD127]) (Figure 6;

Figure 5. Expression levels of the 10 most discriminative genes from blood in classifying patients with sarcoidosis. Relative gene expression levels of microarray data are compared between the internal derivation dataset (University of California, San Francisco [UCSF]: 20 healthy control subjects and 38 patients with sarcoidosis) and the external validation dataset (Oregon: 12 healthy control subjects and 12 patients with sarcoidosis). For the UCSF dataset, all 10 genes were statistically significantly different when comparing sarcoidosis groups with the control group. * $P < 0.05$. † $P < 0.05$ when comparing patients with sarcoidosis with normal lung function to those with low lung function.

Figure 6. Overlap in differentially expressed genes in blood and lung tissue in sarcoidosis. Dots represent the mean log fold change in microarray gene expression of individual genes measured in lung biopsies (OHIO dataset; $n = 6$ patients with sarcoidosis and 6 control subjects) versus whole blood (University of California, San Francisco [UCSF] dataset; $n = 38$ patients with sarcoidosis and 20 control subjects). As described in the online supplement, we identified significant gene expression overlap (red dots) as those genes with an FDR q value of less than 0.1 in both datasets and a greater than 2-fold change in either dataset. Specific genes, such as those that are interferon-inducible or related to the interferon signaling pathway and IL-15, are highlighted.

Table 2). These findings are also concordant with a prior genomics study that identified *IL7* and *IL15* as belonging to the most highly overexpressed gene networks in sarcoidosis lung tissues compared with control subjects (35). Thus, several important molecular pathways were identified in the blood transcriptome, which also appear to be relevant to disease pathogenesis in granulomatous inflammation of the lung.

In addition to examining the overlap of differential gene expression in sarcoid lung tissue compared with blood, we took another approach to examine whether genes associated with progressive fibrotic lung disease in lung tissue were differentially expressed in the peripheral blood of subjects with sarcoidosis. As described in MATERIALS AND METHODS, this dataset did not include healthy control subjects. Therefore, we identified genes differentially expressed in the UCSF blood dataset and compared the overlap in up- or down-regulated genes from the lung biopsies of patients classified as “progressive-fibrotic” by Lockstone

and colleagues (36). Among the top 50 genes up-regulated in the lung tissue dataset (fibrotic versus self-limiting sarcoidosis), nine were significantly differentially expressed in our blood dataset (sarcoidosis versus healthy control) (Table 3). Among the 51 genes down-regulated in the lung tissue dataset, 12 were differentially expressed in our blood dataset (Table 4).

Overlap between Peripheral Blood Transcript Profiles in Sarcoidosis and TB

Given the increasing evidence suggesting exposure to TB as a potential etiologic agent of sarcoidosis, we assessed the qualitative degree of overlap in blood gene expression patterns between sarcoidosis and PTB using the top 50 most discriminating genes from the random forest classifier model based on patients with sarcoidosis versus control subjects. Figure 7A shows the relative gene expression levels depicted in a heat map as well as by principal component analysis. These data reveal overlaps in gene expression patterns between sarcoidosis and PTB and provide a “genomic perspective” of gene regulation in systemic granulomatous inflammation that is shared between *M. tuberculosis* infection and sarcoidosis. We included transcriptional data from other Th1-type driven inflammatory and infectious diseases for comparison. These data reveal that some diseases, such as pediatric lupus, share some features in common with PTB and sarcoidosis.

We also assessed the qualitative degree of gene expression overlap using the 86-gene whole blood transcriptional signature reported by Berry and colleagues to be specific for PTB (29). However, they did not consider sarcoidosis as a disease control in their analyses. Similar to the findings in Figure 7A, the 86-gene, PTB-specific transcriptional signature revealed significant overlap, as shown by heat map and principal component analysis (Figure 7B). Some of the most striking genes depicted in the heat map, such as *DHRS9* and *CCR7*, have indistinguishable expression patterns. Analyses of this set of genes further highlight the similarities between PTB and sarcoidosis expression patterns compared with the other disease conditions.

As an alternative approach to assess the similarities in gene expression patterns in sarcoidosis and PTB, we used the “gene

TABLE 2. THE TWENTY MOST DISCRIMINATIVE GENES FROM BLOOD IN DISTINGUISHING PATIENTS WITH SARCOIDOSIS FROM HEALTHY CONTROL SUBJECTS

Gene Symbol	Gene Name	Entrez Gene ID	Gini*	Log Ratio	Intensity (A value)	FDR q Value
CEACAM1	Carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	634	0.18	0.75	7.60	3.66E-04
CHIC1	Cysteine-rich hydrophobic domain 1	53,344	0.17	−0.52	5.34	5.03E-05
TAP1	Transporter 1, ATP-binding cassette, subfamily B (MDR/TAP)	6,890	0.17	0.37	10.28	5.25E-05
IRF1	Interferon regulatory factor 1	3,659	0.15	0.38	10.04	2.14E-04
ZNF709	Zinc finger protein 709	163,051	0.14	−0.32	4.02	1.00E-04
GBP2	Guanylate binding protein 2, interferon-inducible	2,634	0.13	0.48	11.56	4.71E-05
XPO4	Exportin 4	64,328	0.13	−0.35	6.82	2.14E-04
STX11	Syntaxin 11	8,676	0.13	0.60	8.12	2.73E-05
TTC3	Tetratricopeptide repeat domain 3	7,267	0.12	−0.50	8.27	2.71E-04
CD3G	CD3g molecule, gamma (CD3-TCR complex)	917	0.11	−0.65	9.44	4.38E-04
ZNF512	Zinc finger protein 512	84,450	0.11	−0.33	7.62	1.14E-04
LRRN3	Leucine rich repeat neuronal 3	54,674	0.11	−0.96	5.57	1.59E-04
ZNF662	Zinc finger protein 662	389,114	0.10	−0.47	4.06	1.52E-03
DHRS9	Dehydrogenase/reductase (SDR family) member 9	10,170	0.10	1.08	8.20	2.02E-05
KCNA3	Potassium voltage-gated channel, shaker-related subfamily, member 3	3,738	0.09	−0.66	7.74	2.05E-04
IL7R	Interleukin 7 receptor	3,575	0.09	−0.88	10.43	4.11E-04
ATF3	Activating transcription factor 3	467	0.09	0.40	4.75	1.39E-04
PAQR8	Progesterin and adipoQ receptor family member VIII	85,315	0.09	−0.46	7.51	1.39E-04
GIMAP5	GTPase, IMAP family member 5	55,340	0.09	−0.62	10.31	1.84E-04
CD27	CD27 molecule	939	0.09	−0.74	7.96	4.11E-04

Definition of abbreviation: FDR = false discovery rate.

*The “Gini coefficient” is also known as the “importance statistic” in the random forests algorithm.

TABLE 3. GENES UP-REGULATED IN SARCOIDOSIS LUNG TISSUE ("PROGRESSIVE-FIBROTIC" COMPARED WITH "NODULAR SELF-LIMITING") AND ALSO SIGNIFICANTLY DIFFERENTIALLY EXPRESSED IN UNIVERSITY OF CALIFORNIA, SAN FRANCISCO BLOOD DATASET (FALSE DISCOVERY RATE q VALUE < 0.05)

Symbol	Gene Name	Entrez ID	Probe Set	Log Fold Change in Blood	Average Expression in Blood	FDR q Value for Blood
C1orf38	Chromosome 1 open reading frame 38	9473	9473_at	0.242	9.722	0.005
CLEC7A	C-type lectin domain family 7, member A	64581	64581_at	0.267	9.538	0.048
GBP5	Guanylate binding protein 5	115362	115362_at	1.094	10.185	0.000
IGSF6	Immunoglobulin superfamily, member 6	10261	10261_at	0.196	11.322	0.029
ITGAM	Integrin, alpha M (complement component 3 receptor 3 subunit)	3684	3684_at	0.270	10.915	0.049
PILRA	Paired immunoglobulin-like type 2 receptor alpha	29992	29992_at	0.276	10.310	0.013
PSTPIP2	Proline-serine-threonine phosphatase interacting protein 2	9050	9050_at	0.763	8.249	0.001
SLAMF8	SLAM family member 8	56833	56833_at	0.350	5.565	0.039
SNX10	Sorting nexin 10	29887	29887_at	0.341	8.663	0.011

Definition of abbreviation: FDR = false discovery rate.

module" strategy first described by Chaussabel and colleagues (34), which examines groups of genes that are coordinately expressed, such as genes related to plasma cells, B cells, myeloid lineage, and T cells, as well as other functional categories as previously described (34). Figure E2 shows that gene expression patterns were very similar between sarcoidosis and PTB, with only 4 of 25 modules divergent. Using pathway analysis of the blood transcriptome in sarcoidosis, we found at least nine genes related to Type I and II IFN pathways that overlapped with the blood transcriptional signature reported from subjects with PTB (29) (Figure E1; Table E4). This overlap in blood transcriptional profiles between sarcoidosis and PTB provides evidence that similar activation pathways are engaged in these distinct diseases and may provide insights into biology or serve as genomic markers of disease progression and/or activity.

Although these data suggest that granulomatous lung diseases may share more similarities than differences regarding the blood transcriptional profile, we also explored how blood transcriptional profiles in sarcoidosis and TB may differ. To do this, we used machine learning algorithms to build a classifier with the goal of distinguishing sarcoidosis from PTB. This analysis (Figure 7C) suggests that a number of genes may be divergent between these two diseases, as assessed by the number of genes that are down-regulated in sarcoidosis but not in TB and other Th1-type inflammatory and infectious diseases (most highly divergent gene expression, either up- or down-regulated, was exhibited for genes *GBP6* [guanylate-binding protein6], which is induced by IFN- γ ; *SEPT4* [Septin 4]; *TIMM10* [translocase of inner mitochondrial membrane 10]; and *NOG* [Noggin]). Also of note is the more marked difference in the pattern of gene expression between

sarcoidosis, TB, and another pulmonary granulomatous lung disease from subjects recruited at our center (hypersensitivity pneumonitis) (Figures 7A–7C). Overall, these findings challenge the traditional view that the inflammatory fingerprint of sarcoidosis is confined to diseased organs and suggest that sarcoidosis- and TB-specific signatures can be identified by directed genomic analysis of blood samples despite significant overlap in immunopathology (e.g., shared activation of IFN-related signaling pathways).

DISCUSSION

The goals of this study were (1) to measure the transcriptional signature of peripheral blood from patients with sarcoidosis and determine whether the blood transcriptional profile in sarcoidosis reflects transcriptional abnormalities measured in diseased lungs and (2) to assess overlap in gene expression with a granulomatous infectious disease, PTB, which may play a role in the pathogenesis of sarcoidosis. To this end, we identified a robust transcriptional profile in whole blood from patients with sarcoidosis subjects and identified concordance in gene expression patterns in the blood and diseased lungs, including genes important in the IFN, IL-15, and IL-7 signaling pathways. Finally, we demonstrated significant overlap in gene expression in blood cells in sarcoidosis and active PTB. Our findings indicate that circulating immune cells in sarcoidosis possess significant alterations in genome-wide gene expression supporting a "systemic" inflammatory nature of sarcoidosis. Because we found specific genomic markers in blood that correlate with gene expression patterns in the lung as well as severity of lung function abnormalities, these data justify future

TABLE 4. GENES DOWN-REGULATED IN SARCOIDOSIS LUNG TISSUE ("PROGRESSIVE-FIBROTIC" COMPARED WITH "NODULAR SELF-LIMITING") AND ALSO SIGNIFICANTLY DIFFERENTIALLY EXPRESSED IN UNIVERSITY OF CALIFORNIA, SAN FRANCISCO BLOOD DATASET (FALSE DISCOVERY RATE q VALUE < 0.05)

Symbol	Gene Name	Entrez ID	Probe Set	Log Fold Change in Blood	Average Expression in Blood	FDR q Value for Blood
BMPRI1A	Bone morphogenetic protein receptor, type IA	657	657_at	−0.399	3.322	0.024
CHD6	Chromodomain helicase DNA binding protein 6	84181	84181_at	−0.253	6.071	0.018
EFCAB2	EF-hand calcium binding domain 2	84288	84288_at	0.622	4.510	0.001
FRY	furry homolog (Drosophila)	10129	10129_at	0.404	6.700	0.011
HKR1	HKR1, GLI-Kruppel zinc finger family member	284459	284459_at	−0.340	6.917	0.019
INADL	InaD-like (Drosophila)	10207	10207_at	−0.747	5.896	0.000
ING5	Inhibitor of growth family, member 5	84289	84289_at	−0.181	4.957	0.031
KIAA1328	KIAA1328	57536	57536_at	−0.194	4.963	0.029
OFD1	Oral-facial-digital syndrome 1	8481	8481_at	−0.373	9.249	0.008
PRKAB2	Protein kinase, AMP-activated, beta 2 noncatalytic subunit	5565	5565_at	−0.207	6.267	0.045
ZNF507	Zinc finger protein 507	22847	22847_at	−0.315	6.265	0.008
ZNF512	Zinc finger protein 512	84450	84450_at	−0.354	7.023	0.000

Definition of abbreviation: FDR = false discovery rate.

Figure 7. The distribution of blood gene expression patterns between sarcoidosis and active pulmonary tuberculosis (PTB). Relative gene expression was assessed by generation of heat maps (*left*) and principal component analysis (*right*) as described in MATERIALS AND METHODS. Other infectious and Th1-type inflammatory diseases are included for comparison. (A) Analyses using the most discriminative genes identified by the random forest classifier built to distinguish patients with sarcoidosis from control subjects. (B) Analyses using the PTB-specific gene signature described by Berry and colleagues (29). (C) Analyses using the top 50 most discriminative genes identified by the random forest classifier built to distinguish PTB from sarcoidosis. For heat map figures, genes are listed from top to bottom, and disease cohorts are listed below heat map. From *left to right*: SARC = sarcoidosis (Oregon, n = 12; University of California, San Francisco [UCSF], n = 38); PTB = pulmonary TB (UK, n = 13; UK, n = 21); LTB = latent TB (UK, n = 17; UK, n = 21); PSLE = pediatric systemic lupus erythematosus (n = 49); ASLE = adult systemic lupus erythematosus (n = 28); HP = hypersensitivity pneumonitis (UCSF, n = 6); STAPH = *Staphylococcus aureus* infection (n = 40); STREP = group A *Streptococcus* infection (n = 12).

studies of blood biomarkers of disease progression and phenotype in sarcoidosis, which may be useful in clinical studies.

Although the concept that *M. tuberculosis* may be involved in the development of sarcoidosis is not a new one, more recent studies using biochemical, nucleic acid, and cell-based assays have made important contributions by providing evidence that links mycobacteria to sarcoidosis (23, 25, 28). Collectively, these studies have found nucleic acids and insoluble proteins from mycobacteria in diseased tissues as well as specific T-cell immune responses to mycobacterial antigens independent of length of time of known disease in a reproducible percentage of subjects with sarcoidosis. These published data argue that at least some components of the systemic inflammatory profile in chronic sarcoidosis are stable over time and measurable in peripheral blood. The sarcoidosis gene expression signature we present in this study is similar, but not identical, to TB. These data suggest a role for a TB-like immune response, rather than an infection *per se*, in a certain percentage of individuals who develop sarcoidosis.

Our results raise several questions in light of the available data. Does this genomic data comparison between sarcoidosis and PTB argue further that *M. tuberculosis* plays an important role in the pathogenesis of sarcoidosis? Or have we identified a more general “granulomatous” transcriptional signature that would be observed in many types of granulomatous infectious and noninfectious lung diseases? Although we cannot answer these questions with confidence, transcriptional data presented from subjects with hypersensitivity pneumonitis (Figure 7) argue against a generalized “granulomatous lung disease” signature. Additional diseases, such as chronic berylliosis and histoplasmosis, must be analyzed to address this possibility. Second, can we track disease progression using a genomic transcriptional signature in whole blood? Our data show there are sets of genes that correlate with abnormally low lung function, and validation through longitudinal studies is necessary to address their utility in monitoring disease progression and severity. Finally, can we use the peripheral blood transcriptional signature to guide therapeutic decisions by tracking response to therapy? This concept was elegantly demonstrated by Berry and colleagues in monitoring treatment in patients with TB (29) and would be particularly helpful in managing patients with sarcoidosis who have chronic disease because we now have a larger armamentarium of potential therapies (e.g., methotrexate, TNF- α inhibitors, mycophenolate mofetil, thalidomide, etc.) and few tools for monitoring response.

Although sarcoidosis is traditionally thought to be a disease in which the pathogenic immune cells are “compartmentalized” in the diseased tissue or organ (4, 48), our data identify a set of genes that are concordantly regulated in blood and lung and that plausibly contribute to the development or persistence of granulomatous inflammation (IFN- α , IFN- γ , IL-15, and IL-7). We speculate that the breadth and sensitivity conferred by newer genomic approaches to characterize systemic inflammation (34) may explain our findings of overlap between blood and lung signatures of specific inflammatory pathways. The circulating immune cells expressing these genes, which are concordant in blood and lung, may identify cells that have recently migrated to the diseased organ and are contributing to ongoing inflammation. Immunophenotyping and functional analysis of these cells may enhance our understanding of the immune biology of sarcoidosis and elucidate novel immune subsets that could be targeted in new treatment approaches. In summary, we provide evidence for a robust peripheral blood transcriptional signature in sarcoidosis, which raises questions about how we should think of disease “compartmentalization” and could be leveraged to help us understand the immunobiology of this disease, provide markers of disease severity, and identify novel immune cells that contribute to organ inflammation.

Author Disclosure: None of the authors has a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

Acknowledgment: The authors thank the members of the Interstitial Lung Disease Clinic and specifically Paul Wolters, Harold Chapman, and Sally McLaughlin as well as the participants in the study.

References

1. Siltzbach LE, James DG, Neville E, Turiaf J, Battesti JP, Sharma OP, Hosoda Y, Mikami R, Oda M. Course and prognosis of sarcoidosis around the world. *Am J Med* 1974;57:847–852.
2. James DG. Immunology of sarcoidosis. *Lancet* 1966;2:633–635.
3. Sones M, Israel HL. Altered immunologic reactions in sarcoidosis. *Ann Intern Med* 1954;40:260–268.
4. Hudspeth BN, Flint KC, Geraint-James D, Brostoff J, Johnson NM. Lack of immune deficiency in sarcoidosis: compartmentalisation of the immune response. *Thorax* 1987;42:250–255.
5. Hirschhorn K, Schreiberman RR, Bach FH, Siltzbach LE. In-vitro studies of lymphocytes from patients with sarcoidosis and lymphoproliferative diseases. *Lancet* 1964;2:842–843.
6. Sharma OP, James DG, Fox RA. A correlation of in vivo delayed-type hypersensitivity with in vitro lymphocyte transformation in sarcoidosis. *Chest* 1971;60:35–37.
7. Kataria YP, Sagone AL, LoBuglio AG, Bromberg PA. In vitro observations on sarcoid lymphocytes and their correlation with cutaneous energy and clinical severity of disease. *Am Rev Respir Dis* 1973;108:767–776.
8. Robinson BW, McLemore TL, Crystal RG. Gamma interferon is spontaneously released by alveolar macrophages and lung T lymphocytes in patients with pulmonary sarcoidosis. *J Clin Invest* 1985;75:1488–1495.
9. Greene CM, Meachery G, Taggart CC, Rooney CP, Coakley R, O'Neill SJ, McElvaney NG. Role of IL-18 in CD4+ T lymphocyte activation in sarcoidosis. *J Immunol* 2000;165:4718–4724.
10. Muller-Quernheim J, Saltini C, Sondermeyer P, Crystal RG. Compartmentalized activation of the interleukin 2 gene by lung T lymphocytes in active pulmonary sarcoidosis. *J Immunol* 1986;137:3475–3483.
11. Crystal RG, Roberts WC, Hunninghake GW, Gadek JE, Fulmer JD, Line BR. Pulmonary sarcoidosis: a disease characterized and perpetuated by activated lung T-lymphocytes. *Ann Intern Med* 1981;94:73–94.
12. Hunninghake GW, Gadek JE, Young RC Jr, Kawanami O, Ferrans VJ, Crystal RG. Maintenance of granuloma formation in pulmonary sarcoidosis by T lymphocytes within the lung. *N Engl J Med* 1980;302:594–598.
13. Hunninghake GW, Crystal RG. Pulmonary sarcoidosis: a disorder mediated by excess helper T-lymphocyte activity at sites of disease activity. *N Engl J Med* 1981;305:429–434.
14. Daniele RP, Dauber JH, Rossman MD. Immunologic abnormalities in sarcoidosis. *Ann Intern Med* 1980;92:406–416.
15. Costabel U, Bross KJ, Ruhle KH, Lohr GW, Matthys H. Ia-like antigens on T-cells and their subpopulations in pulmonary sarcoidosis and in hypersensitivity pneumonitis: analysis of bronchoalveolar and blood lymphocytes. *Am Rev Respir Dis* 1985;131:337–342.
16. Saltini C, Spurzem JR, Lee JJ, Pinkston P, Crystal RG. Spontaneous release of interleukin 2 by lung T lymphocytes in active pulmonary sarcoidosis is primarily from the leu3+dr+ T cell subset. *J Clin Invest* 1986;77:1962–1970.
17. Rossi GA, Sacco O, Cosulich E, Risso A, Balbi B, Ravazzoni C. Helper T-lymphocytes in pulmonary sarcoidosis: functional analysis of a lung T-cell subpopulation in patients with active disease. *Am Rev Respir Dis* 1986;133:1086–1090.
18. Saltini C, Hemler ME, Crystal RG. T lymphocytes compartmentalized on the epithelial surface of the lower respiratory tract express the very late activation antigen complex via-1. *Clin Immunol Immunopathol* 1988;46:221–233.
19. Pinkston P, Bitterman PB, Crystal RG. Spontaneous release of interleukin-2 by lung T lymphocytes in active pulmonary sarcoidosis. *N Engl J Med* 1983;308:793–800.
20. Hunninghake GW, Bedell GN, Zavala DC, Monick M, Brady M. Role of interleukin-2 release by lung T-cells in active pulmonary sarcoidosis. *Am Rev Respir Dis* 1983;128:634–638.

21. Hunninghake GW, Crystal RG. Mechanisms of hypergammaglobulinemia in pulmonary sarcoidosis: site of increased antibody production and role of T lymphocytes. *J Clin Invest* 1981;67:86–92.
22. Barnard J, Rose C, Newman L, Canner M, Martyny J, McCammon C, Bresnitz E, Rossman M, Thompson B, Rybicki B, *et al.* Job and industry classifications associated with sarcoidosis in a case-control etiologic study of sarcoidosis (access). *J Occup Environ Med* 2005; 47:226–234.
23. Chen ES, Moller DR. Etiology of sarcoidosis. *Clin Chest Med* 2008;29: 365–377 (vii).
24. Newman LS, Rose CS, Bresnitz EA, Rossman MD, Barnard J, Frederick M, Terrin ML, Weinberger SE, Moller DR, McLennan G, *et al.* A case control etiologic study of sarcoidosis: environmental and occupational risk factors. *Am J Respir Crit Care Med* 2004;170:1324–1330.
25. Drake WP, Dhasan MS, Nadaf M, Shepherd BE, Vadivelu S, Hajizadeh R, Newman LS, Kalams SA. Cellular recognition of mycobacterium tuberculosis ESAT-6 and KATG peptides in systemic sarcoidosis. *Infect Immun* 2007;75:527–530.
26. Gupta D, Agarwal R, Aggarwal AN, Jindal SK. Molecular evidence for the role of mycobacteria in sarcoidosis: a meta-analysis. *Eur Respir J* 2007;30:508–516.
27. Klemen H, Husain AN, Cagle PT, Garrity ER, Popper HH. Mycobacterial DNA in recurrent sarcoidosis in the transplanted lung: a PCR-based study on four cases. *Virchows Arch* 2000;436:365–369.
28. Song Z, Marzilli L, Greenlee BM, Chen ES, Silver RF, Askin FB, Teirstein AS, Zhang Y, Cotter RJ, Moller DR. Mycobacterial catalase-peroxidase is a tissue antigen and target of the adaptive immune response in systemic sarcoidosis. *J Exp Med* 2005;201:755–767.
29. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, Wilkinson KA, Banchereau R, Skinner J, Wilkinson RJ, *et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 2010;466:973–977.
30. Woodruff PG, Solberg OD, Snyder-Cappione JE, Hou L, Nguyen C, Chi J, Koth LL. Whole blood gene expression analysis identifies sarcoidosis-specific markers including decreased IL7 receptor expression. *Am J Respir Crit Care Med* 2010;181:A3981.
31. Statement on sarcoidosis. Joint statement of the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of Sarcoidosis and other Granulomatous Disorders (WASOG) adopted by the ATS Board of Directors and by the ERS Executive Committee, February 1999. *Am J Respir Crit Care Med* 1999;160:736–755.
32. Rosenbaum JT, Pasadhika S, Crouser ED, Choi D, Harrington CA, Lewis JA, Austin CR, Diebel TN, Vance EE, Brazier RM, *et al.* Hypothesis: sarcoidosis is a STAT1-mediated disease. *Clin Immunol* 2009;132:174–183.
33. Winterbauer RH, Belic N, Moores KD. Clinical interpretation of bilateral hilar adenopathy. *Ann Intern Med* 1973;78:65–71.
34. Chaussabel D, Quinn C, Shen J, Patel P, Glaser C, Baldwin N, Stichweh D, Blankenship D, Li L, Munagala I, *et al.* A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity* 2008;29:150–164.
35. Crouser ED, Culver DA, Knox KS, Julian MW, Shao G, Abraham S, Liyanarachchi S, Macre JE, Wewers MD, Gavrillin MA, *et al.* Gene expression profiling identifies MMP-12 and ADAMDEC1 as potential pathogenic mediators of pulmonary sarcoidosis. *Am J Respir Crit Care Med* 2009;179:929–938.
36. Lockstone HE, Sanderson S, Kulakova N, Baban D, Leonard A, Kok WL, McGowan S, McMichael AJ, Ho LP. Gene set analysis of lung samples provides insight into pathogenesis of progressive, fibrotic pulmonary sarcoidosis. *Am J Respir Crit Care Med* 2010;181:1367–1375.
37. Diaz-Uriarte R, Alvarez de Andres S. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics* 2006;7:3.
38. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc, B* 2005;67:310–320.
39. Dabney AR. Classification of microarrays to nearest centroids. *Bioinformatics* 2005;21:4148–4154.
40. Liaw A, Wiener M. Classification and regression by randomforest. *R News* 2002;2:18–22.
41. Woodruff PG, Boushey HA, Dolganov GM, Barker CS, Yang YH, Donnelly S, Ellwanger A, Sidhu SS, Dao-Pick TP, Pantoja C, *et al.* Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids. *Proc Natl Acad Sci USA* 2007;104:15858–15863.
42. Dolganov GM, Woodruff PG, Novikov AA, Zhang Y, Ferrando RE, Szubin R, Fahy JV. A novel method of gene transcript profiling in airway biopsy homogenates reveals increased expression of a Na⁺-K⁺-Cl⁻ cotransporter (nkcc1) in asthmatic subjects. *Genome Res* 2001;11:1473–1483.
43. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;5:R80.
44. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser A Stat Soc* 2005;57:289–300.
45. Wahlstrom J, Katchar K, Wigzell H, Olerup O, Eklund A, Grunewald J. Analysis of intracellular cytokines in CD4⁺ and CD8⁺ lung and blood T cells in sarcoidosis. *Am J Respir Crit Care Med* 2001;163: 115–121.
46. Konishi K, Moller DR, Saltini C, Kirby M, Crystal RG. Spontaneous expression of the interleukin 2 receptor gene and presence of functional interleukin 2 receptors on T lymphocytes in the blood of individuals with active pulmonary sarcoidosis. *J Clin Invest* 1988;82:775–781.
47. Schwarz MI, King TE. Interstitial lung disease. BC Decker; Hamilton, Ontario, Canada: 1998.
48. Hunninghake GW, Fulmer JD, Young RC Jr, Gadek JE, Crystal RG. Localization of the immune response in sarcoidosis. *Am Rev Respir Dis* 1979;120:49–57.