**Cathepsin S, a new serum biomarker of sarcoidosis discovered by transcriptome analysis of alveolar macrophages**

**Hiroyuki Tanaka**, Etsuro Yamaguchi, Nobuhiro Asai, Toyoharu Yokoi, Masaki Nishimura, Haruhisa Nakao, Masashi Yoneda, Yoshinori Ohtsuka, Satoshi Konno, Noritaka Yamada

1 Division of Respiratory Medicine and Allergology, Department of Internal Medicine, School of Medicine, Aichi Medical University, Nagakute, Aichi, Japan; 2 Department of Infection Control and Prevention, School of Medicine, Aichi Medical University, Nagakute, Aichi, Japan; 3 Department of Pathology, Nagoya Ekisaikai Hospital, Nagoya, Aichi, Japan; 4 Division of Hepatology and Pancreatology, Department of Internal Medicine, School of Medicine, Aichi Medical University, Nagakute, Aichi, Japan; 5 Department of Medicine, Hokkaido Chuo Rosai Hospital, Iwamizawa, Japan; 6 First Department of Medicine, Hokkaido University School of Medicine, Sapporo, Japan; 7 Department of Respiratory Medicine, National Hospital Organization, Higashinagoya Hospital, Nagoya, Aichi, Japan

**Abstract.** Background: Development of reliable new biomarkers remains crucial to improve diagnosis and assessing disease activity in sarcoidosis. The objective of this study was to seek such markers from the gene expression signature of alveolar macrophages by transcriptome analysis. **Methods:** Pooled RNA extracted from alveolar macrophages from patients with active sarcoidosis and control patients was subjected to transcriptome analysis using microarrays. **Results:** Among 12 genes with ratios higher than that of a housekeeping gene, we selected CTSS for scrutinizing protein expression in serum because of the feasibility of the protein assay. CTSS concentrations were significantly increased in sarcoidosis compared with not only controls but also all the other lung diseases. Receiver operating characteristics curve for sarcoidosis and parenchymal lung diseases revealed an area under the curve of 0.800 (95% confidence interval, 0.751-0.850; p=1.4 x 10^-18) with 70% sensitivity and 78% specificity at a CTSS concentration of 15.5 ng/ml. **Conclusions:** CTSS has the potential to be a useful biomarker in sarcoidosis. (Sarcoidosis Vasculitis and Diffuse Lung Diseases 2018; 35: 00-00)

**Key words:** biomarkers, cathepsins, macrophages, sarcoidosis, transcriptome

**Introduction**

Sarcoidosis is a widespread multisystem disease that preferentially involves the lungs, intra-thoracic lymph nodes, eyes, and skin. The gold standard of diagnosis includes pathological findings compatible with sarcoidosis and exclusion of other infectious and non-infectious granulomatous diseases along with the presence of compatible clinical features (1). The presence of multiple typical clinical manifestations
alone such as bilateral hilar lymph node enlargement, uveitis, and skin rash strongly suggests sarcoidosis but insufficient for confident diagnosis. In such cases, positive results for serum biomarkers with high sensitivity and specificity are a strongly supportive of the diagnosis. Since the discovery of angiotensin converting enzyme (ACE) by Lieberman (2), several but not many markers have been added to the list of candidate markers, such as soluble IL-2 receptor (3), Krebs von den Lungen 6 (4), tryptase (5), and amyloid A (6). Recently, omics studies such as of the transcriptome of blood (7,8) and the proteomes of alveolar macrophages (9) and serum (10,11) have been conducted. However, translation of the results from those omics studies to the development of biomarkers useful in clinical practice has been unsuccessful mainly because test specimens are hard to obtain or assays require complicated laboratory works (7-9).

Since no biomarkers have been validated as gold surrogates for diagnosis and the monitoring of disease course of sarcoidosis, such new markers are sorely needed. The current study attempted to seek such markers by conducting comparative transcriptome analysis between patients with sarcoidosis and controls. As specimens for screening by transcriptome analysis, we selected alveolar macrophages as cells that are easily obtained by bronchoalveolar lavage (BAL) and are believed to reflect immune responses at the site of disease.

Materials and Methods

Study Subjects

This study included not only healthy volunteers and patients with sarcoidosis but also a wide variety of disease controls in order to fairly assess the specificity of potential markers (Table 1). All subjects were Japanese.

Healthy volunteers without any prior major illness or any symptoms at the time of blood sampling served as controls. They all were never-smokers and were not included if they reported regular medication use.

Patients with sarcoidosis had evidence of non-caseating epithelioid cell granulomas in at least one organ and compatible clinical features in at least two organs such as bilateral hilar and/or mediastinal lymph node enlargement with or without lung parenchymal infiltrates, eye lesions, and skin lesions, and without evidence of mycobacterial, fungal, or parasitic infection as described in the American Thoracic Society/European Respiratory Society/World Association of Sarcoidosis and Other Granulomatous Disorders publications (1). None had a history of exposure to organic or inorganic materials known to cause lung diseases. No patients were taking systemic corticosteroids when serum was first tested for CTSS concentration. The presence or absence of active lesions in each organ was assessed by inspection of skin visually and palpation, auscultation of lungs and heart, chest roentgenography, spirometry, electrocardiography, echocardiography, routine laboratory tests including leukocyte count, hemoglobin concentration, hepatic enzyme activity, calcium ion concentration, and ACE activity, and ophthalmic examination by specialists. Since apparent absence of active lesions assessed by these noninvasive routine clinical examinations does not necessarily mean truly inactive state, the current study included eight patients who had been previously diagnosed but had no clinical signs of affected organs at the time of serum collection. The number of patients in Scadding stage

<table>
<thead>
<tr>
<th>Table 1. Demographic data of study subjects*</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>sarcoidosis</td>
</tr>
<tr>
<td>Interstitial pneumonia</td>
</tr>
<tr>
<td>Pneumococcosis</td>
</tr>
<tr>
<td>Pulmonary mycobacteriosis</td>
</tr>
</tbody>
</table>

* Data are median [range]. Definitions of abbreviations: FEV1, forced expiratory volume in 1 second; FVC, forced vital capacity; DLco, diffusing capacity of lung for carbon monoxide; VA, volume of alveolar gas; BAL, bronchoalveolar lavage
Cathepsin S as a new biomarker of sarcoidosis

0, stage I, stage II, and stage III were 16 (15%), 56 (52%), 31 (29%), and 4 (4%), respectively. Clinical courses of sarcoidosis were judged by changes in the number of affected organs or apparent changes in the severity of lesions clinically assessed as described above.

Patients with idiopathic interstitial pneumonia had either a usual interstitial pneumonia pattern (n=14) or a nonspecific interstitial pneumonia pattern (n=12) on high-resolution computed tomography according to the ATS/ERS consensus classification (12) or a published document (13) with no evidence of underlying diseases judging from occupational history, physical examination, and serological tests for collagen vascular diseases or other vasculitides. No patients were taking systemic corticosteroids when sera were collected.

Pneumoconiosis was diagnosed based on longstanding occupational exposure to silica or coal, and discrete nodular shadows of variable sizes in predominantly upper lung areas on chest roentgenography or computed tomography. According to the management classification by the Ministry of Health, Labor and Welfare of Japan, 50 patients were classified into type 2 (a number of opacities smaller than 1 cm in diameter [ILO classification: p, q, r, s, t, u]), 50 into type 3 (numerous opacities smaller than 1 cm in diameter), and 50 into type 4 (nodular shadows larger than 1 cm in diameter).

Diagnosis of pulmonary mycobacteriosis was made based on compatible radiological findings and demonstration of mycobacteria in sputum and/or bronchoalveolar lavage fluid. Forty-six patients had tuberculosis and 30 had nontuberculous mycobacteriosis.

**Microarray**

BAL was conducted in three nonsmoking patients with active sarcoidosis with biopsy evidence and four nonsmoking controls who were suspected to have lung cancer. BAL was done in the middle lobe or lingular segments for patients with sarcoidosis, or in segments unaffected by pulmonary lesions for patients suspected of lung cancer. All were women. Since most lymphocytes of BAL cells are cluster of differentiation (CD) 2+ T cells (14) and the proportion of granulocytes was less than 3% in all the subjects studied, a cellular fraction of BAL cells depleted of CD2+ cells by microbeads coated with anti-CD2 monoclonal antibody (MicroBeads, Miltenyi Biotec GmbH, Bergisch Gladbach, Germany) was used as alveolar macrophages. Total RNA from this cell fraction was extracted using BioRobot EZ1 (Qiagen GmbH, Hilden, Germany) and EZ1 RNA Cell Mini Kit (Qiagen). Equal amount of total RNA from each subject was pooled for the patient group or the control group, and 0.4 μg of RNA from each group was subjected to microarray analysis using CodeLink Human Whole Genome Bioarray (Applied Microarrays, Tempe, AZ) and CodeLink Expression Bioarray System (Applied Microarrays). Array slides were hybridized with biotin-labelled cRNA, stained with streptavidin-Cy5, washed according to the manufacturer’s protocol, and scanned by arrayWoRx (Applied Precision, Issaquah, WA).

**Measurement of CTSS Protein and Immunostaining**

Sera were obtained at the time of annual health check-up in healthy volunteers, at the time of initial visit or at the oldest time point in patients with sarcoidosis, and at variable time points in patients with other diseases. Concentrations of CTSS in sera diluted 100-fold with assay diluent were measured using a Human Total Cathepsin S DuoSet (R&D Systems, Minneapolis, MN). Specific immunostaining for human CTSS was performed on formalin-fixed paraffin-embedded slides prepared from a mediastinal lymph node in a patient with sarcoidosis using goat anti-CTSS antibody (M-19, sc-6505; Santa Cruz Biotechnology, Santa Cruz, CA), and peroxidase-labelled anti-goat immunoglobulin (Histofine Simple Stain MAX-PO [G]; Nichirei Bioscience, Tokyo, Japan).

**Statistical Analysis**

Difference between two study groups was assessed by the Mann-Whitney test followed by Bonferroni’s correction to counteract multiple comparisons. Receiver operating characteristics curve was used to determine sensitivity and specificity for discriminating sarcoidosis from control subjects or other lung diseases. The Youden index was used to determine the cut-off point. The Jonckheere-Terpstra test was used to assess trend in CTSS concentrations across patient groups ranked by the number of...
affected organs. The Wilcoxon signed-rank test was used to compare paired concentrations of CTSS during the clinical course of sarcoidosis. These statistical analyses were performed using SPSS Statistics 23.0 (SPSS, Chicago, IL). P values less than 0.05 were considered statistically significant. Sample size estimation was calculated by the PS program (Power and Sample Size Calculation version 3.1.2, http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize).

The ethics committee of the Department of Medicine at Aichi Medical University approved all study protocols (No 260, No 2009-18, No 13-141), and all study subjects provided informed consent prior to participation.

**Results**

Among approximately 34,000 array probes including expressed sequence tags, those expressing good quality signals were selected. Ratios of normalized signal intensity for individual genes in the sarcoidosis sample to that in the control sample were calculated. Twelve genes showed ratios higher than 6.7 which was the ratio for glyceraldehyde-3-phosphate dehydrogenase, a housekeeping gene used as the reference gene (Table 2). The standard deviation of ratio of all genes on microarray was approximately 2.0 assuming the ratio as continuous variable. If the difference of ratio means between sarcoidosis and controls was set as more than 5.7 (control=1.0 and sarcoidosis >6.7, Table 2), we need only 3 experimental subjects and 3 control subjects to be able to reject the null hypothesis that the means of the experimental and control groups are equal with power 0.8 and the type I error probability 0.05. Thus, this power calculation justified the actual sample size (3 controls and 4 patients) in the present study. Since enzyme-linked immunosorbent assay (ELISA) kits for the target proteins in serum were not available for the first- and second-ranked genes (LRAP and ZNF101, respectively, in Table 2), we selected the third-ranked gene, CTSS, for which there was a commercially available ELISA kit.

We measured serum concentrations of CTSS in not only sarcoidosis patients and healthy controls, but also in patients with several pulmonary diseases. They were significantly increased not only in sarcoidosis, but also in other diseases except for pulmonary mycobacteriosis compared with those in healthy controls (Figure 1). Although ages differed significantly between controls and other diseases (p<0.001), no significant correlation was identified between CTSS concentrations and ages in control subjects (p=-0.71, p=0.507).

Table 2. Ratios of probe intensity for individual genes in a pooled sarcoidosis sample to that in a pooled control sample

<table>
<thead>
<tr>
<th>Ratio</th>
<th>Description (NCBI/UniGene database)</th>
</tr>
</thead>
<tbody>
<tr>
<td>31.2</td>
<td>leukocyte-derived arginine aminopeptidase (LRAP)</td>
</tr>
<tr>
<td>24.3</td>
<td>cDNA DKFZp570I0164 (ZNF101)</td>
</tr>
<tr>
<td>21.7</td>
<td>cathepsin S (CTSS)</td>
</tr>
<tr>
<td>11.9</td>
<td>28206323prime NIH MGC 7 cDNA clone</td>
</tr>
<tr>
<td>10.8</td>
<td>UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase</td>
</tr>
<tr>
<td>9.1</td>
<td>cathepsin D (CTSD)</td>
</tr>
<tr>
<td>7.5</td>
<td>cDNA DKFZp686P21116</td>
</tr>
<tr>
<td>7.2</td>
<td>tyR83b061 NCI CGAP Kid11 cDNA clone</td>
</tr>
<tr>
<td>7.1</td>
<td>secretoglobin, family 3A, member 1 (SCGB3A1)</td>
</tr>
<tr>
<td>7.0</td>
<td>hepatitis C virus core-binding protein 6 (HCBP6)</td>
</tr>
<tr>
<td>6.9</td>
<td>602533729F1 NIH MGC 15 cDNA clone</td>
</tr>
<tr>
<td>6.9</td>
<td>insulin-like growth factor 1 (somatomedin C) (IGF1)</td>
</tr>
<tr>
<td>6.7</td>
<td>glyceraldehyde-3-phosphate dehydrogenase (GAPDH)</td>
</tr>
</tbody>
</table>

**Fig. 1.** Serum CTSS concentrations.

Data are presented in box-and-whisker plots with ends of the whiskers representing the lowest datum still within the 1.5 interquartile range (IQR) under the lower quartile and the highest datum still within 1.5 IQR above the upper quartile. Circles represent outliers between 1.5 and 3.0 IQR under the lower quartile or above the upper quartile. Asterisks represent extreme values below 3.0 IQR under the lower quartile or above 3.0 IQR over the upper quartile. The horizontal dotted line indicates the 95th percentile value of controls.

* Significantly higher than controls. P values for SA, IP, and PC were 1.3X10^-30, 8.6X10^-8, and 4.0X10^-31, respectively. † Significantly higher than the other pulmonary diseases. P values for IP, PC, and PM were 6.7X10^-6, 7.8X10^-8, and 2.1X10^-6, respectively.

Definitions for abbreviations: CO, control; SA, sarcoidosis; IP, interstitial pneumonia; PC, pneumoconiosis; PM, pulmonary mycobacteriosis.
The median value in sarcoidosis was the highest (17.9 ng/ml) and was significantly increased compared with all other diseases (Figure 1). Ninety-nine of the 107 (93%) patients with sarcoidosis exceeded the upper 95th percentile value (11.7 ng/ml) of controls. In receiver operating characteristics curve analysis for control and sarcoidosis, area under the curve was 0.985 (95% CI, 0.972–0.997; p=1.8 x 10⁻⁵). Controls and patients with sarcoidosis were most effectively discriminated at a CTSS concentration of 11.9 ng/ml with 93% sensitivity and 97% specificity. When interstitial pneumonia, pneumoconiosis, and pulmonary mycobacteriosis were combined into a group of parenchymal lung diseases which is usually subject to differentiation from sarcoidosis, sarcoidosis and the combined group were most effectively discriminated at a CTSS concentration of 15.5 ng/ml with 70% sensitivity and 78% specificity (Figure 2).

An increasing trend in CTSS concentrations was seen among patients with sarcoidosis as a greater number of organs are affected, and this trend was significant by the Jonckheere-Terpstra test (p=0.008). CTSS concentrations were measured at a given reference date, mostly at the initial visit, and reevaluated at a later date in a proportion of patients with sarcoidosis (n=40). No significant change was seen in nine patients with unchanged activity (median [range] number of affected organs at the reference date and date of reevaluation: 1 [1–4] and 1 [1–4], respectively) (Figure 3A). Meanwhile, significant decreases were seen in 19 patients with spontaneous improvement (2 [1–5] and 1 [0–4], respectively) (Figure 3B) and in seven patients following the administration of systemic corticosteroids (3 [2–6] and 2 [1–3], respectively) (Figure 3C). Opposite changes were observed in five patients with disease progression (1 [1–4] and 1 [1–5], respectively) (Figure 3D).

Immunohistochemical staining with anti-CTSS antibody for lymph nodes obtained from nine patients with sarcoidosis was positive for epithelioid cells in sarcoid granulomas (Figure 4).

Discussion

We have conducted a transcriptome study of alveolar macrophages for the screening of serum biomarkers that are potentially adoptable for clini-
cal use in sarcoidosis. We found that CTSS had the potential to be such a biomarker, which has not been reported in previous proteomic or transcriptomic studies (7-9,11). We used samples of pooled RNA that were expressed by alveolar macrophages in order to save the number of arrays required and evaluated the relative expression of each RNA by calculating the ratio of the signal intensity in sarcoidosis patients to that in the control subjects. Although this method is not frequently used (15), it proved to be effective in finding a new serum biomarker in sarcoidosis.

The sensitivity for sarcoidosis was higher for the CTSS concentrations than those for the other markers reported to date (6,10,16,17). The source of CTSS in serum was thought to be alveolar macrophages and epithelioid cells in the granulomas, since CTSS was expressed by these cells and the serum concentrations tended to rise in parallel with the number of affected organs, which roughly reflected the total granuloma load. Similar to ACE as a classic marker of sarcoidosis, CTSS could also be a marker of disease activity, because its concentrations varied in parallel with both the natural clinical course and in response to corticosteroid treatment.

The specificity of CTSS concentrations in terms of discriminating other lung diseases was modest, as these were elevated in a proportion of patients with other parenchymal lung diseases. This fact suggested that CTSS partially reflected inflammation in general, like the other serum markers of sarcoidosis (3,18,19,20). Nevertheless, the CTSS concentrations in sarcoidosis were significantly higher than those in all the other lung diseases examined. This fact enabled establishment of an appropriate cut-off concentration of CTSS, which is potentially useful for differentiating sarcoidosis from other parenchymal lung diseases that occasionally exhibit lung shadows that are similar to those in sarcoidosis.

CTSS is expressed in B cells, macrophages, and dendritic cells and is required for invariant chain (Ii) degradation and antigen processing (21,22). Ii is a peptide within antigen-presenting cells and covers the peptide-binding groove of the major histocompatibility complex (MHC) II molecules until it encounters a foreign peptide in the lysosomal compartment of the cells (7). In the presence of interferon-γ, CTSS becomes the main enzyme that can cleave Ii, leaving class II-associated Ii peptides, which are then displaced by HLA-DM in the presence of antigens. This allows antigen peptides to bind to class II molecules and to be transported to the cell membrane for presentation (22,23). CTSS is supposedly involved in the pathogenic mechanism of sarcoidosis through these processes.

Acknowledgments
We are grateful to Ms. Kamiya and Ms. Takagi for their excellent technical assistance. This study was supported in part by a grant of the Strategic Research Foundation Grant-aided Project for Private Universities from the Ministry of Education, Culture, Sports, Science, and Technology, Japan (MEXT), 2011-2015 (S1101027) and by a grant for Rare Lung Diseases (pulmonary alveolar proteinosis, congenital interstitial lung disease) from the Ministry of Health Labor and Welfare, Japan.

References
7. Bloom CI, Graham CM, Berry MP, et al. Transcriptional blood signa-