# Identification of Lineage Markers for T Cell Immune Dysregulation in Sarcoidosis Using Single-Cell RNA-seq

Akihiro Nomura, Hideo Matsuda*
Graduate School of Information Science and Technology, Osaka University, Suita, Osaka, Japan.

* Corresponding author. Tel.: +81-6-6879-4390; email: matsuda@ist.osaka-u.ac.jp
Manuscript submitted February 18, 2022; accepted June 5, 2022.
doi: 10.17706/ijbbb.2022.12.4.63-70

**Abstract:** This paper proposes a method for the identification of the expressed markers from single-cell RNA-seq data in sarcoidosis patients. This method classifies T cells into several subtypes and infers lineages that undergo a transition process to each subtype as the disease progresses. It also estimates each ongoing continuous gene expression and identifies lineages and markers by analyzing differences within and between lineages. Using the method, we have revealed how naive T cell changes as the disease progresses in sarcoid patients and successfully detected potential markers in sarcoidosis patients, which was not detected by existing cluster-based methods.

**Key words:** Single-cell RNA-seq, trajectory inference, T cell immune dysregulation, sarcoidosis.

## 1. Introduction

Sarcoidosis is a systemic inflammatory disease characterized by the infiltration of immune cells into granulomas. Previous gene expression studies using heterogeneous cell mixtures lack insight into cell-type-specific immunomodulatory dysfunction. The first single-cell RNA sequencing study of sarcoidosis in control peripheral immune cells was reported [1]. Genes that are differentially expressed in 18 cell types have been identified and bioinformatically evaluated for functional and pathway enrichment. Their results revealed sustained activation of circulating classical monocytes followed by upregulation of trafficking molecules. In addition, the sarcoidosis T cell subset showed a pattern of dysregulation. But this study is performed to cluster cells, but trajectory analysis of the T cell differentiation process is not performed. Therefore, no sarcoid-related information on the differentiation of T cells into their respective subtypes is available.

In this paper, we propose a novel method for analyzing the differentiation process of T cells in sarcoid patients. The study was to identify markers that are expressed in T cell subtypes during differentiation. We performed a trajectory analysis to analyze how T cell subtypes change as the disease progresses in sarcoidosis. Expression analysis between and within the trajectories was performed to clarify the marker expressed in gamma delta T cells, which is one of the subtypes of T cells in the process of T cell differentiation.

Our method consists of the following five steps: 1) clustering of cells; 2) extracting the CD4+ T cells from all types of cells to analyze the T cell differentiation process; 3) dimensional reduction with UMAP; 4) performing trajectory inference to identify lineages of transitional processes along with T cell differentiation progression; and 5) marker detection by differential expression analysis between lineages (Fig. 1). By using this method, we can identify marker candidates that are differentially expressed as specific lineages progress toward sarcoidosis differentiation.

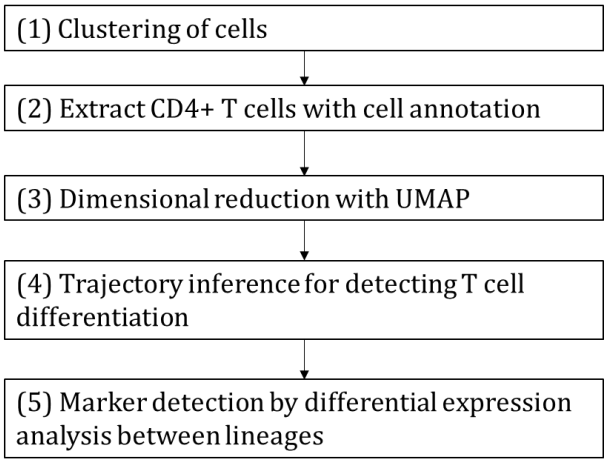



Fig. 1. Workflow of our method.

## 2. Marker Analysis of Lineages of T cell differentiation

### 2.1. The Dataset in Previous Research

We used scRNA-seq data to analyze Sarcoidosis from the Gene Expression Omnibus repository (GSE132338). The data were derived with scRNA-seq of 35 patients and 13 healthy controls cells from patient-derived PBMC. Briefly, assessment consisted of a one-time visit, where, after providing informed consent, subjects donated a blood sample and completed surveys of demographics, medication usage, and clinical histories. In this trajectory analysis, we used the cell type annotation information in the previous research paper [1]. Using a standard analysis pipeline with Seurat [2], cells were clustered by similar gene expressions. To characterize cell-type-specific transcriptomes, they generated scRNA-seq profiles of over 100,000 PBMC. Following quality control measures, they identified 18 cell clusters (data not shown), all of which could be assigned cell identities by canonical markers. 13 cell types contained at least 1,000 cells. Each subject contributed an average of 1,280 cells that could be identified. To analyze T cell differentiation, we extracted CD4 memory T cells, CD4 naive T cells, early T effector T cells, gamma delta T cells, and regulatory T cells. We performed PCA analysis and UMAP dimensional reduction using 20 dimensions (Fig. 2). As shown in Fig. 1, five large clusters (CD4 memory T cells, CD4 naive T, early T effector T cells, gamma delta T cells, and regulatory T cells) were observed.

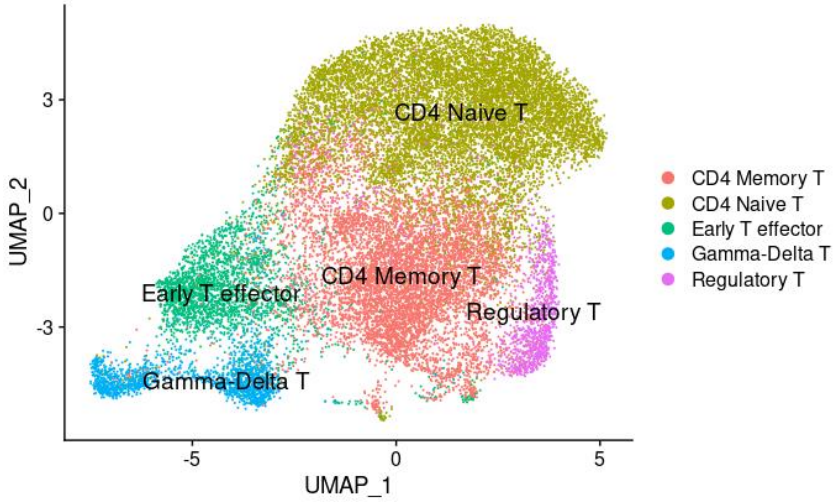
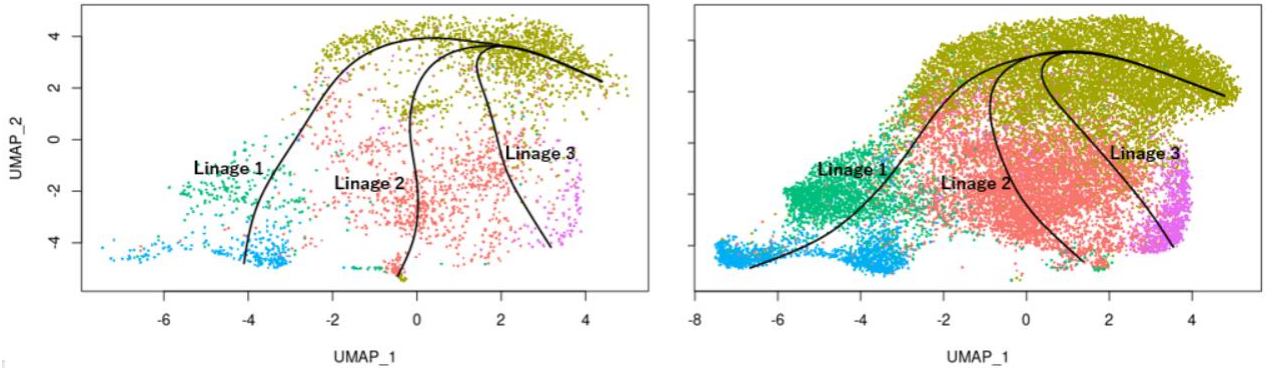



Fig. 2. UMAP plots of T cells with cell types.

## 2.2. Lineage Analysis

We conducted trajectory inference with Slingshot [3]. In the inference, we set cluster CD4 naive T cell as a starting cluster and detected three lineages (Fig. 2). These lineages denote the lineage that progressed from CD4 naive T cell to gamma delta T cell (lineage 1), CD4 memory T cell (lineage 2), and regulatory T cell (lineage 3). The differentiation process of each T-cell subtype from CD4 naive cells is successfully visualized by lineage analysis. Interestingly, different populations of healthy controls and sarcoid patients were observed in gamma delta T cells. In addition, the tip direction was different in healthy controls (Fig. 3 (a)) and sarcoidosis (Fig. 3(b)) in linage 1 due to their different cell populations.




(a) (b)

Fig. 3. Healthy control and patient lineage detected by trajectory inference.

## 2.3. Identifying Markers of Between-Lineage Differential Expression

Given the lineages identified in Section 2.2, we estimated lineage-specific gene expression. To compute continuous gene expression across the clusters of the lineage, we employed tradeSeq [4] to infer functions for gene expression along pseudotime for each lineage using generalized additive models. We performed several statistical tests on expression within and between the lineages described in Section 2.2. To discover markers of the T cell states transitions of T cell differentiation, we assessed differential expression between the starting and ending points along lineages [5]. In addition, to detect genes that express exclusively in gamma delta T cells, we performed a difference Waldstat test between healthy control and sarcoidosis gamma delta T cells (linage 1) and regulatory T cells (linage 3).

We used a Wald test to assess the null hypothesis that the average expression at the starting point (CD4 naive T cell) is equal to the average expression at the endpoints (gamma delta T cell, CD4 memory T cell, and regulatory T cell). Then we subtracted the regulatory T cells (linage 3) Waldstat score from the gamma delta T cells (linage 1) Waldstat score. Table 1 shows the markers detected by the resulting scores. The marker NKG7 and GNLY were reported in a previous research paper [1] and CCL, CST7, GZMC, and TRDC are found by a marker analysis described in Section 2.4.

Interestingly, we detected new genes CTSW, PLAC8, ID2, and ANXA1 by the method. Each gene described above continuous changes in the expression of markers along lineage progression is shown in Fig. 4. The four genes show a large difference in statistical scores between lineages 1 and 3 in healthy patients from the statistical scores between lineages 1 and 3 in sarcoid patients. As for gene ID2, the difference in expression between the lineages 1 and 3 of sarcoid patients appeared to be smaller than that of healthy controls, but the Waldstat score difference between the lineages of sarcoid patients and healthy controls was large when judged from a statistical perspective.

Table 1. Markers Detected by Difference between Lineages 1 and 3

| | difference of Waldstat 1vs3 | logFC1vs3 |
|---|---|---|
| CCL5 | 203.60 | 4.88 |
| NKG7 | 115.73 | 6.08 |
| GNLY | 113.99 | 6.29 |
| CST7 | 111.71 | 2.57 |
| CTSW | 99.62 | 3.77 |
| PLAC8 | 91.74 | 3.42 |
| GZMA | 84.04 | 2.73 |
| ID2 | 83.04 | 3.29 |
| TRDC | 80.39 | 5.26 |
| ANXA1 | 78.82 | 2.11 |

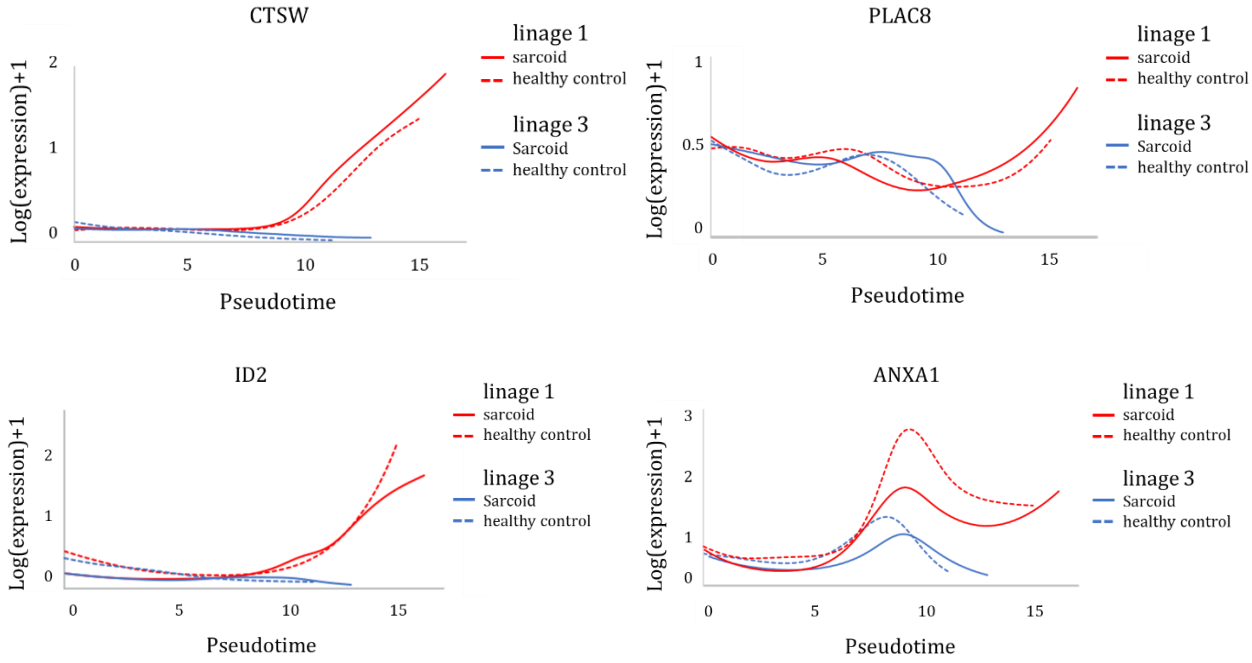



Fig. 4. Continuous changes in the expression of markers along with lineage progression.

## 2.4. Comparison with Cluster-Based Marker Analysis

As described in Section 2.3, we also used Seurat (FindMarkers with the options: only.pos=TRUE) to compare the cluster-based markers in sarcoidosis gamma delta T cells and sarcoidosis CD4+ naive T cells. Table 2 shows the markers found by the analysis. Compared to genes found in Section 2.3 between-Lineage differential expression analysis, six of the top ten genes (GNLY, NKG7, CCL5, GZMA, TRDC, CST7) were common. Of the genes found in the analysis, KLRB1 was reported in a previous study, while TRGC1 and DUSP2 were not found in the differential lineage analysis of Section 2.3 and were only found in the analysis. PLAC8, ID2, CTSW, and ANXA1 were not found in the cluster-based analysis with Seurat. This suggests that our detection of markers along lineage progression can provide a different means for understanding sarcoidosis progression mechanisms.

Table 2. Markers Detected by Seurat Marker Analysis

| | avg_logFC |
|---|---|
| GNLY | 3.349 |
| NKG7 | 3.33 |
| CCL5 | 2.77 |
| KLRB1 | 2.74 |
| GZMA | 2.27 |
| TRDC | 2.22 |
| CST7 | 2.09 |
| TRGC1 | 2.08 |
| DUSP2 | 1.87 |

## 3. Discussion

Sarcoidosis is an enigmatic disease with a well-established genetic predisposition and immune dysregulation [6], [7], but no other etiology. For this reason, it is important to understand the immune dysregulation inherent in sarcoidosis. To date, researchers are working to develop an effective method for identifying gene expression markers by scRNA-seq that goes beyond bulk RNA-seq. In this paper, we performed a lineage-based differential expression analysis to identify markers that are differentially expressed as a particular lineage progresses toward one of the T cells, gamma delta T cells. Gamma delta T cells are known to play a myriad of roles in pathogen clearance, wound healing, autoimmunity, and cancer, primarily through the production of soluble mediators [8].

Two of six markers detected commonly by the differential lineage analysis and cluster-based analysis, NKG7 and GNLY were already reported in previous research [1]. The other 4 genes like CCL5, CST7, GZMA, and TRDC are known as follows. CCL5 is associated with pulmonary sarcoidosis [9]. CST7 was observed among the most highly upregulated sarcoidosis [10]. GZMA has been highly expressed in highly effector (TEMRA) memory T cells [11]. TEMRA memory T cell is a subset of effector memory T cells re-expresses CD45RA (termed TEMRA) after antigenic stimulation with unknown molecular characteristics and functions. CD4 TEMRA cells have been implicated in protective immunity against pathogens such as dengue virus (DENV) [12]. TRDC encodes the T cell receptor delta constant region, one component of the gamma delta T cell receptor which is essential for the development of gamma delta T cells [13].

Among the genes found only in the cluster-based marker analysis by Seurat are TRGC1 and DUSP2. TRGC1 (Constant region of T cell receptor (TR) gamma chain) participates in antigen recognition. Gamma-delta TRs recognize a variety of self and foreign non-peptide antigens [14]. Upon antigen recognition induces rapid, innate-like immune responses involved in pathogen clearance and tissue repair [15], [16]. DUSP2 (dual specificity phosphatase 2) is also known as phosphatase of activated cells 1, PAC1 acts as an immune checkpoint in T cell antitumor immunity. PAC1 is selectively upregulated in exhausted tumor-infiltrating lymphocytes and is associated with poor prognosis of patients with cancer [17].

Furthermore, in this paper, we performed differential expression analysis between specific lineages and identified new genes differentially expressed in gamma delta T cells. Namely, CTSW, PLAC8, ID2, and ANXA1 were detected by differential values in each of the lines comparing lineage 1 (gamma delta cells) and lineage 3 (regulatory T cells). CTSW is highly expressed in TEMRA as well as GZMA [11]. PLAC8 was a commonly expressed gene in pulmonary sarcoidosis in comparison to transcribed blood in pulmonary tuberculosis and pulmonary sarcoidosis [10]. ID2 is a transcription factor involved in the development of lung ILC cells (Innate

lymphoid cells), and ILCs are involved in lung function, barrier integrity, and respiratory tissue remodeling after lung injury [18]. These results suggest that the phylogenetic differential analysis method is effective in detecting genes related to autoimmune diseases like sarcoidosis that cannot be detected by cluster-based analysis. ANXA1 Lipocortin-1 (also known as annexin-1) is upregulated by corticosteroids and plays a prominent part in many of their anti-inflammatory actions [19] and is one mediator of endogenous anti-inflammation, affording regulation of leukocyte trafficking and activation in many contexts [20]. ANXA1 peptide-mimetic, given therapeutically, ameliorated both signs of inflammation and fibrosis [20].

## 4. Conclusion

In this paper, we propose a new method to detect markers that differ in a lineage of T cell differentiation between sarcoidosis cases and healthy controls. In other words, we proposed a new method to detect the difference between genes that vary when lineages migrate from naive t cells to gamma delta T cells and genes that vary when lineages migrate from naive T cells to regulatory T cells. Our method can analyze continuous state transitions based on single-cell transcriptome data and successfully detects potentially variable genes such as CTSW, PLAC8, ID2, and ANXA1. This is something that could not be detected by conventional cluster-based variation gene analysis tools or ordinary phylogenetic analysis. Differential analysis of lineage-specific variable genes will be an effective tool for understanding the process of T-cell state change during sarcoid disease progression. In the future, we aim to develop tools that can identify more biomarkers and targets as potential lineage trajectories. With the improvement of the trajectory inference method, we aim to develop a tool that can identify more potential lineages.

### Conflict of Interest

The authors declare no conflict of interest.

### Author Contributions

Akihiro Nomura and Hideo Matsuda conducted the research, analyzed the data, and wrote the paper; both authors approved the final version of the manuscript.

### Acknowledgment

This work was supported in part by JSPS KAKENHI Grant Numbers JP18H04124, JP19K22894, JP20H04947, and JP21K19827 Japan.

### References

[1] Garman, L., Pelikan, R. C., Rasmussen, A., Lareau, C. A., Savoy, K. A., Deshmukh, U. S., *et al.* (2020). Single cell transcriptomics implicate novel monocyte and T cell immune dysregulation in sarcoidosis. *Frontiers in Immunology, 11*, 567342.

[2] Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., *et al*. (2019). Comprehensive integration of single-cell data. *Cell*, *177(7)*, 1888-1902.

[3] Street, K., Risso, D., Fletcher, R., Das, D., Ngai, J., Yosef, N., *et al.* (2018). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, *19(1)*, 477.

[4] Van den Berge, K., Roux de Bézieux, H., Street, K., Saelens, W., Cannoodt, R., Saeys, Y., *et al.* (2020). Trajectory-based differential expression analysis for single-cell sequencing data. *Nature Communications*, *11(1)*, 1201.

[5] Murayama, K., & Matsuda, H. (2021). Detecting lineage-specific marker genes for tumor evolution based on single cell transcriptome. *International Journal of Bioscience, Biochemistry and Bioinformatics, 11(3)*,

50-57.

[6] Moller, D. R., Rybicki, B. A., Hamzeh, N. Y., Montgomery, C. G., Chen, E. S., Drake, W., *et al.* (2017). Genetic, immunologic, and environmental basis of sarcoidosis. *Annals of the American Thoracic Society, 14*, S429–S436.

[7] Bello, G. A., Adrianto, I., Dumancas, G. G., Levin, A. M., Iannuzzi, M. C., Rybicki, B. A., *et al.* (2015). Role of NOD2 pathway genes in sarcoidosis cases with clinical characteristics of blau syndrome. *American Journal of Respiratory Critical Care Medicine, 192(9)*, 1133–1135.

[8] Carding, S. R., & Egan, P. J. (2002). Gamma delta T cells: Functional plasticity and heterogeneity. *Nature Review Immunology, 2(5)*, 336-345.

[9] Hashemi, N., Weigt, S., Xue, Y., Palchevskiy, V., Huang, C., Zisman, D., *et al.* (2009). Immune response CC chemokines, CCL5 and CCL2 are associated with pulmonary sarcoidosis. *American Journal of Respiratory and Critical Care Medicine, 179*, A4011.

[10] Bloom, C. I., Graham, C. M., Berry, M. P., Rozakeas, F., Redford, P. S., Wang, Y., *et al*. (2013). Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers, *Plos One, 8(8)*, e70630.

[11] Cano-Gamez, E., Soskic, B., Roumeliotis, T. I., So, E., Smyth, D. J., Baldrighi, M., *et al*. (2020). Single-cell transcriptomics identifies an effectorness gradient shaping the response of CD4+ T cells to cytokines. *Nature Communications, 11*, 1801.

[12] Tian, Y., Babor, M., Lane, J., *et al*. (2017). Unique phenotypes and clonal expansions of human CD4 effector memory T cells re-expressing CD45RA. *Nature Communications, 8*, 1473

[13] Petersen, B., Kammerer, R., Frenzel, A., Hassel, P., Dau, T. H., Becker, R., *et al*. (2021). Generation and first characterization of TRDC-knockout pigs lacking γδ T cells. *Scientific Reports, 11*, 14965

[14] Lefranc, M. P. (2014). Immunoglobulin and T cell receptor genes: IMGT (®) and the Birth and rise of immunoinformatics. *Frontiers in Immunology, 5*, *22*.

[15] Nielsen, M. M., Witherden, D. A., & Havran, W. L. (2017). γδ T cells in homeostasis and host defence of epithelial barrier tissues. *Nature Reviews Immunology, 17(12)*, 733-745.

[16] Vantourout, P., & Hayday, A. (2013). Six-of-the-best: Unique contributions of γδ T cells to immunology. *Nature Reviews Immunology, 13 (2)*, 88-100.

[17] Lu, D., Liu, L., Sun, Y., Song, J., Yin, Q., Zhang, G., *et al*. (2020). The phosphatase PAC1 acts as a T cell suppressor and attenuates host antitumor immunity. *Nature Immunology, 21(3)*, 287-297.

[18] Monticelli, L. A., Sonnenberg, G. F., Abt, M. C., Alenghat, T., Ziegler, C. G., Doering, T. A., *et al.* (2011). Innate lymphoid cells promote lung-tissue homeostasis after infection with influenza virus. *Nature Immunology 12*, 1045–1054.

[19] Mikuniya, T., Nagai, S., Shigematsu, M., Hamada, K., Hoshino, Y., Mio, T., *et al.* (2001). Lipocortin-1 (also known as annexin-1) is upregulated by corticosteroids and plays a prominent part in many of their anti-inflammatory actions. *Sarcoidosis Vasculitis, and Diffuse Lung Diseases, 18(3)*, 272-278.

[20] Damazo, A. S., Sampaio, A. L., Nakata, C. M., Flower, R. J., Perretti, M., & Oliani, S. M. (2011). Endogenous annexin A1 counter-regulates bleomycin-induced lung fibrosis. *BMC Immunology, 12*, 59.

**Akihiro Nomura** received his B.Sc. degree in science from Tokyo Institute of Technology, Japan in 2002, and received his M.E. degree in 2004. His research interests include genomic data analysis, gene regulatory networks, and gene expression analysis in the field of autoimmune disease.

**Hideo Matsuda** received his B.Sc. degree in physics from Kobe University, Japan in 1982, and he received his M.E. and Ph.D. degrees in computer science from Kobe University, Japan in 1984 and 1987, respectively. He has been a professor in the Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University since 2002. His research interests include genomic data analysis, gene regulatory networks, and gene expression analysis.