

Gene Network Inference from Single-Cell Omics Data and Domain Knowledge for Constructing COVID-19-Specific *ICAM1*-Associated Pathways

Mitsuhiro Odaka^{1,2,3,*}, Morgan Magnin^{2,3,+}, and Katsumi Inoue^{1,2,3,+}

¹Department of Informatics, The Graduate University for Advanced Studies, SOKENDAI, Tokyo, Japan

²Principles of Informatics Research Division, National Institute of Informatics, Tokyo, Japan

³Laboratoire des Sciences du Numérique de Nantes, École Centrale de Nantes, Nantes, France

⁴Japan Society for the Promotion of Science, Tokyo, Japan

*Corresponding author: Mitsuhiro Odaka; E-mail: odaka[at]nii[dot]ac[dot]jp; TEL: +818057061107

+These authors contributed equally to this work.

Supplementary Information

This document contains supplementary information on single-cell omics data analysis and pathway mapping. The single-cell data analysis extracts differentially expressed genes (DEGs) and differentially coexpressed genes (DCGs).

Differentially expressed gene analysis

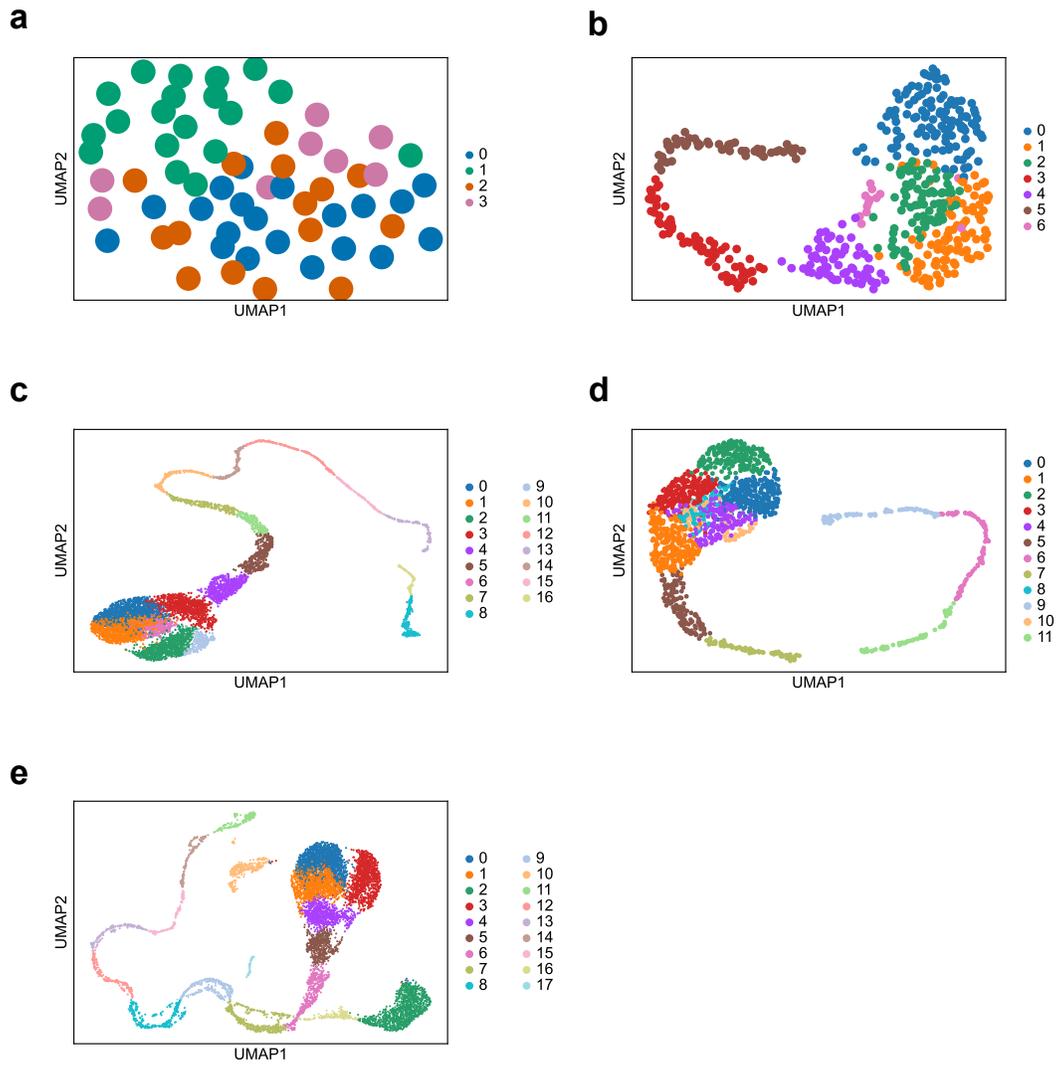
Supplementary Figs. S1, S2, and S3 depict the results of the extracted DEGs. Supplementary Fig. S1 presents the genewise clustering results. Principal component analysis (PCA) and uniform manifold approximation and projection (UMAP) embedded data points onto the two-dimensional space. Louvain method, which does not require manual input of cluster counts, classified these data points into any number of clusters. Fig. S2 shows heatmaps for each cluster based on the signal intensity of gene expression. Each cluster contains the top five genes only. The dendrogram on the right side of the heatmaps gives the order among the clusters. Fig. S3 lists the top 10 genes with significant differences in expression changes for each cluster by the Wilcoxon rank-sum test. The Wilcoxon rank-sum test evaluates the difference in representative values between two data groups without assuming normality by substituting the observed values into the rank order. Calculating the rank-sum score of genes detected the significant difference, sorting the genes within each cluster in descending order.

Differentially coexpressed gene analysis

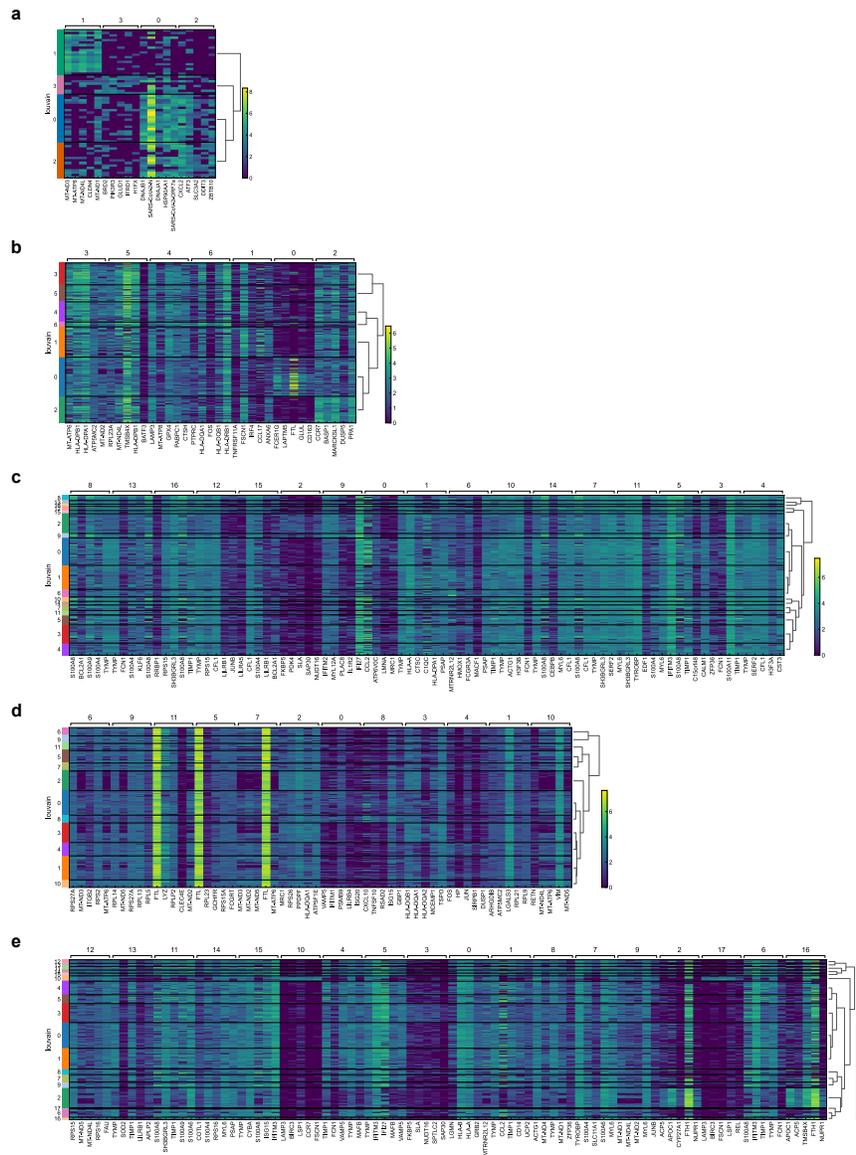
Supplementary Figs. S4, S5, and S6 depict the results of the extracted DCGs. Analyzing the data for DCG extraction follows the above methods to obtain DEGs.

Pathway Mapping

Pathway mapping discovers which subpathways within existing signaling pathways reflect the activity of a group of genes varying and coexpressed in a disease-specific manner in the observed gene expression data that one has (Supplementary Fig. S7).



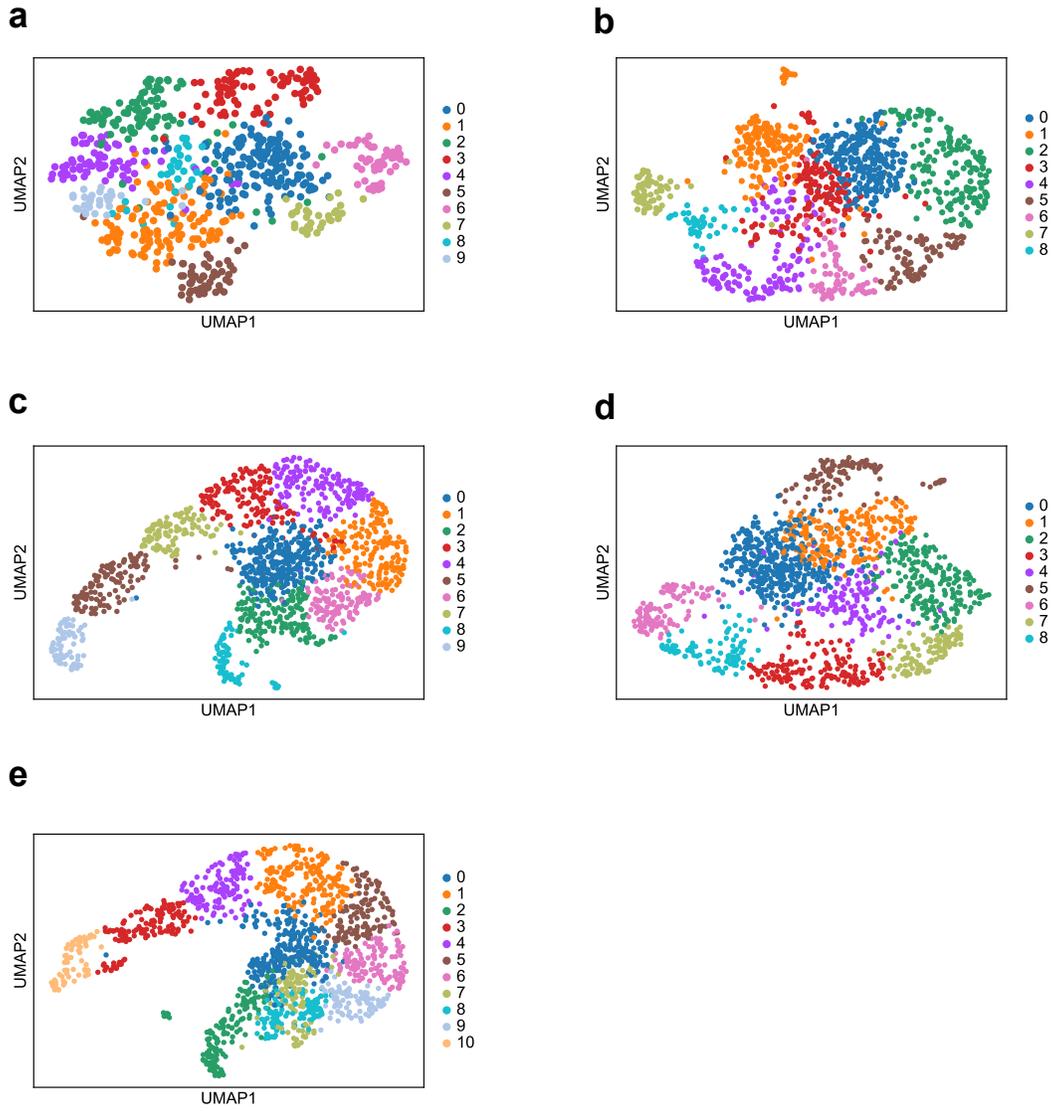
Supplementary Fig. 1. Genewise clustered data points within embedded latent space. The Euclidean distance measured the distance between clusters. Calculating the log fold change (the magnitude of differential expression) yielded a list of the differential expression genes (DEGs), of which expression level significantly increased or decreased. See also [10.6084/m9.figshare.17263814](https://doi.org/10.6084/m9.figshare.17263814).



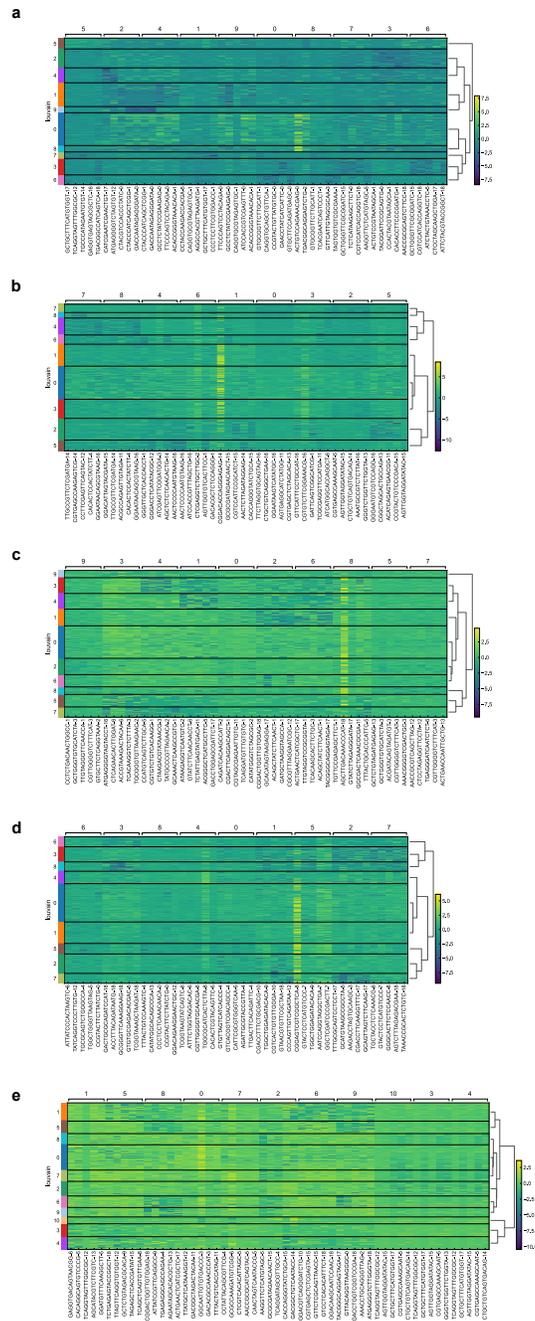
Supplementary Fig. 2. Top five differential expressed genes (DEGs) for each cluster. See also [10.6084/m9.figshare.17263841](https://doi.org/10.6084/m9.figshare.17263841).



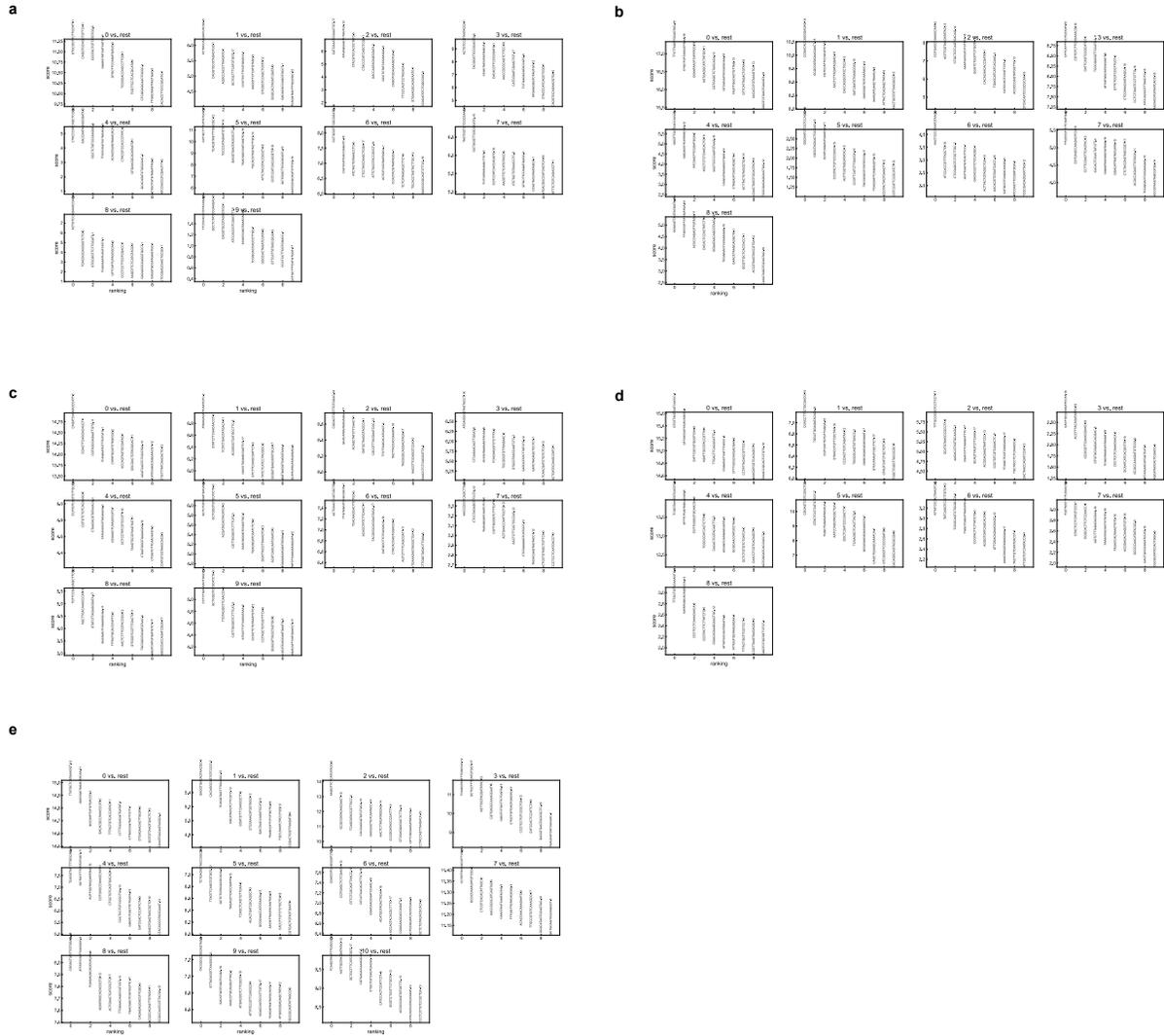
Supplementary Fig. 3. The Wilcoxon rank-sum test determined clusters' rank-ordered differentially expressed genes (DEGs). Each cluster of each subfigure contains the top 10 genes' names in descending order from left to right, with rank-sum scores assigned to the vertical axes. See also [10.6084/m9.figshare.17263877](https://doi.org/10.6084/m9.figshare.17263877).



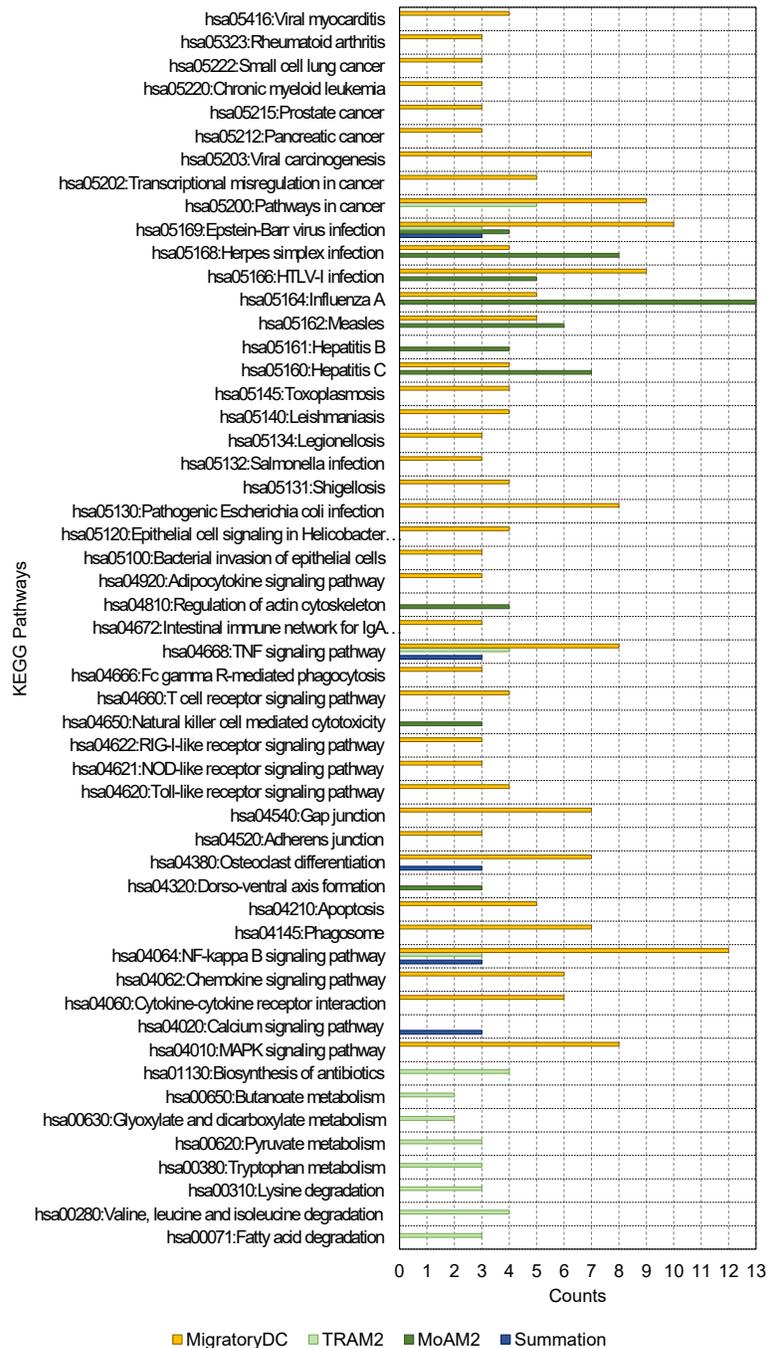
Supplementary Fig. 4. Cellwise clustered data points within embedded latent space. Pearson's correlation coefficient measures the distance between clusters. See also [10.6084/m9.figshare.17263889](https://doi.org/10.6084/m9.figshare.17263889).



Supplementary Fig. 5. Top five cell IDs for each cell cluster with a similar gene expression pattern. See also [10.6084/m9.figshare.17263892](https://doi.org/10.6084/m9.figshare.17263892).



Supplementary Fig. 6. Differentially coexpressed genes (DCGs) for each cluster. Each cluster of each subfigure contains the top 10 cells' IDs in descending order from left to right, with rank-sum scores assigned to the vertical axes. See also [10.6084/m9.figshare.17263898](https://doi.org/10.6084/m9.figshare.17263898).



Supplementary Fig. 7. Querying the gene lists of directed graphical models provided the activated Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways for each cell type. The bars colored yellow, lime green, olive, and navy blue stand for the count of genes that appeared on the KEGG pathways for migratory dendritic cells (DCs), tissue-resident alveolar macrophages (TRAM2), monocyte-derived alveolar macrophages (MoAM2), and all cells. See also [10.6084/m9.figshare.17263907](https://doi.org/10.6084/m9.figshare.17263907).