



Inferring microenvironmental regulation of gene expression from single-cell RNA sequencing data using scMLnet with an application to COVID-19

Jinyu Cheng, Ji Zhang, Zhongdao Wu and Xiaoqiang Sun

Corresponding author: Xiaoqiang Sun, Zhong-Shan School of Medicine, Sun Yat-Sen University, Guangzhou 510080, China.
Tel.: +86-020-87330128; Fax: +86-020-87331209; E-mail: sunxq6@mail.sysu.edu.cn; xiaoqiangsun88@gmail.com

Abstract

Inferring how gene expression in a cell is influenced by cellular microenvironment is of great importance yet challenging. In this study, we present a single-cell RNA-sequencing data based multilayer network method (scMLnet) that models not only functional intercellular communications but also intracellular gene regulatory networks (<https://github.com/SunXQlab/scMLnet>). scMLnet was applied to a scRNA-seq dataset of COVID-19 patients to decipher the microenvironmental regulation of expression of SARS-CoV-2 receptor ACE2 that has been reported to be correlated with inflammatory cytokines and COVID-19 severity. The predicted elevation of ACE2 by extracellular cytokines EGF, IFN- γ or TNF- α were experimentally validated in human lung cells and the related signaling pathway were verified to be significantly activated during SARS-COV-2 infection. Our study provided a new approach to uncover inter-/intra-cellular signaling mechanisms of gene expression and revealed microenvironmental regulators of ACE2 expression, which may facilitate designing anti-cytokine therapies or targeted therapies for controlling COVID-19 infection. In addition, we summarized and compared different methods of scRNA-seq based inter-/intra-cellular signaling network inference for facilitating new methodology development and applications.

Key words: scMLnet; single-cell RNA-seq; multilayer network; cellular microenvironment; SARS-CoV-2

Introduction

The gene expression and cellular functioning are not determined by the intracellular substance only but are also dependent on extracellular signals [1]. Various types of cells and cytokines exist in the extracellular microenvironment [2], which constitute a heterogeneous environment for cell growth, migration and function. The microenvironmental signals can be transmitted into the intracellular molecular regulatory network through multiple layers of signal propagation including receptors, signaling

molecules, transcription factors and target genes [3]. Therefore, inferring intercellular communications and intracellular signaling networks would help to deepen our understanding of gene-expression regulation and cellular fate determination [4, 5].

The emerging single-cell sequencing technology provides us with an unprecedented opportunity to deconvolute microenvironment composition and to predict cell-cell communications. At the resolution of single cell, cell-specific biological effects

Jingyu Cheng is a graduate student at Zhong-Shan School of Medicine, Sun Yat-Sen University.

Ji Zhang, PhD, is an associate professor at Department of Neurosurgery, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-Sen University Cancer Center.

Zhongdao Wu, PhD, is a professor at Zhong-Shan School of Medicine, Sun Yat-Sen University. His research interests include bioinformatics and omics-based infection and immune research.

Xiaoqiang Sun, PhD, is an associate professor at Zhong-Shan School of Medicine, Sun Yat-Sen University. His research interests include computational systems biology and bioinformatics.

Submitted: 21 July 2020; Received (in revised form): 18 September 2020

can be captured to depict cell heterogeneity [6]. The single-cell RNA-sequencing (scRNA-seq) transcriptomic data can be used to analyze interacting cell types and cell type-specific gene expressions, so as to elaborate the microenvironment-mediated intercellular and intracellular signaling pathways [7].

In recent years, several methods or tools have been developed to infer cell–cell communications from scRNA-seq data. For example, SoptSC [8] predicts pathway-mediated cell–cell communications by defining ligand–receptor (L–R) signaling probabilities based on gene expression of specific pathway genes in single cells. SingleCellSignalR [9] provides a curated LR database and calculates a novel regularized score to assess the confidence in predicted LR interactions by controlling false positives. A more comprehensive LR database CellPhoneDB [10] has been created by taking into account the subunit architecture of both ligands and receptors and a statistical framework was developed to predict enriched cellular interactions between two cell types from single-cell transcriptomics data. Furthermore, considering important spatial distance between cells, SpaOTsc [11] infers cell–cell communications by ‘optimally transporting’ signal senders to target signal receivers via structured optimal transport to recover spatial properties of scRNA-seq data in the aid of the spatial reference. More recently, CellChat [12] has been developed to quantitatively infer and analyze intercellular communication networks from scRNA-seq data by means of network analysis, pattern recognition and manifold learning approaches.

However, the above methods or tools mainly focus on intercellular communications but fall short on modeling how gene expression of a cell is influenced by interacting cells. In fact, the signaling events involved in cell–cell communications include not only intercellular signals but also intracellular signaling pathways and transcriptional regulations [3]. To predict ligand–target gene links between interacting cells, NicheNet [13] integrates prior knowledge on ligand–receptor, signaling and gene regulatory networks into weighted networks and used a network propagation method to calculate a regulatory potential score between all pairs of potential active ligands and predefined target genes. Although NicheNet can be applied to both bulk and single cell expression data, the prior model of ligand–target regulatory potential depends mainly on prior network information rather than expression relationships in specific cells. Therefore, context-dependent multilayer, inter- and intra-cellular signaling networks are required to be constructed for functionally understanding cell–cell communications by more fully taking advantage of the single cell gene expressions.

In this study, we developed a scRNA-seq data-driven multilayer network method (scMLnet) that models not only intercellular communications but also intracellular gene regulatory networks. scMLnet integrates intercellular pathways (ligand–receptor interactions) and intracellular subnetworks (receptor–TF pathways and TF–target gene interactions) based on cell-type specific gene expression, prior network information and statistical inference. scMLnet could be used to predict microenvironmental regulators of gene expression within a cell using scRNA-seq data and to understand functional roles of intercellular communications.

We then applied scMLnet to a scRNA-seq dataset of COVID-19 patients to investigate microenvironmental regulation of expression of SARS-CoV-2 receptor gene ACE2 (angiotensin-converting enzyme 2), which has been reported to be correlated with inflammatory cytokines and COVID-19 severity but the underlying mechanisms remain poorly understood. To this end,

we collect published scRNA-seq data of COVID-19 patients to analyze the cellular compositions and to construct both intercellular and intracellular signaling networks. Specifically, we construct subnetworks connecting to ACE2 and integrate them into a multi-cellular multilayer network. Based on the network analysis, several key regulators of ACE2 expression were revealed. Subsequently, using independent bulk transcriptomic data from COVID-19 patient samples, several pathways in the multilayer network were verified to be activated during SARS-CoV-2 infection. At last but not least, we experimentally validate the predicted regulation of ACE2 by cytokines including EGF, IFN- γ and TNF- α . Our study also has therapeutic relevance for intervention of COVID-19. We also analyzed the sensitivity/robustness of our method to the variation of the cell clustering parameters. In addition, we summarized and compared different methods of cell–cell communication inference to facilitate tools’ applications and new methodology development.

Methods

scRNA-seq data collection and quality control

The scRNA-seq data of bronchoalveolar lavage fluid (BALF) in COVID-19 patients were collected from the Gene Expression Omnibus (GEO) database with accession number GSE145926 [14]. The dataset contained nine samples of COVID-19 patients, including three patients in moderate stage and six patients in severe stage. In addition, we included four samples of BALF in healthy controls (GSE145926 [14] and GSM3660650 [15]) for accurate and comprehensive cell clustering and cell-type annotation. The scRNA-seq data were subject to quality control to obtain high-quality data for the following-up analysis. We perform quality control using the following criteria (gene number between 200 and 6000, UMI count >1000 and mitochondrial gene percentage <0.1 [14]). All the 13 samples of scRNA-seq data were used for cell-clustering analysis, and then the expression matrix of cells with annotated identity in the nine patient samples were used as input of scMLnet to infer multilayer networks of ACE2 regulation during SARS-CoV-2 infection.

Cell type identification

The integration, normalization, standardization, dimensionality reduction and tSNE visualization of the expression matrix were performed using R package Seurat v3.1.2 [16].

The ‘NormalizeData’ function of Seurat was used to normalize the expression matrix of single cells. The method ‘Log-Normalize’ was used to multiply the expression matrix by 10 000, which was then divided by the size of the total library, so that different cells could be comparable. The top 2000 highly differentially expressed genes for principal component analysis were identified using the ‘FindVariableGenes’ function (vst method as default). The ‘ScaleData’ were used to scale and center the expression levels of highly variable genes, excluding the influence of mitochondrial genes and the total number of molecules detected within a cell. Finally, tSNE method was used for visualization. The ‘FindClusters’ function was used for single-cell clustering, and the cell types were identified according to the expression levels of marker genes from the original studies. Appropriate parameters were selected according to the original study [14].

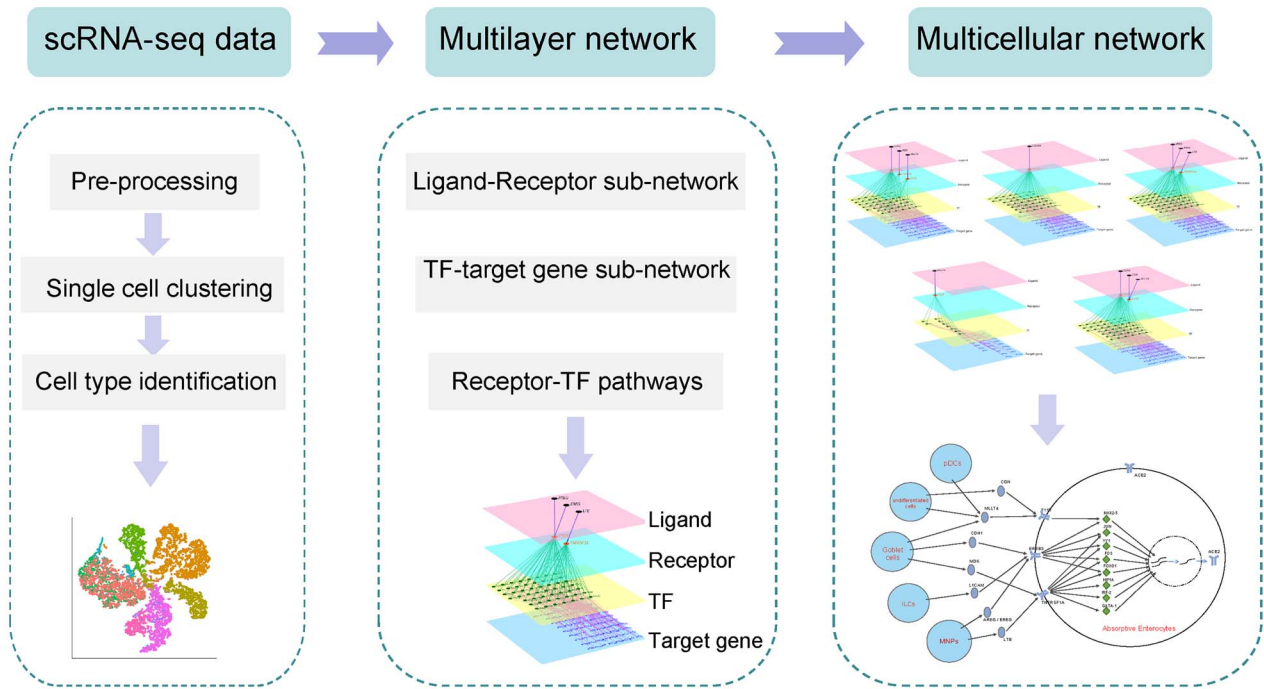


Figure 1. A schematic diagram of scMLnet. (i) We first processed the RNA-seq data and performed clustering analysis to identify cell types according to specific marker genes. (ii) We then constructed multilayer network by integrating intercellular pathways (ligand-receptor interactions) and intracellular subnetworks (receptor-TF pathways and TF-target gene interactions) based on cell-type specific gene expression, prior network information and statistical inference (Fisher's exact test and correlation). (iii) A multicellular network was constructed by connecting different microenvironmental cells to the target cells via combining multilayer networks, to elucidate microenvironment-mediated regulation of gene expression.

Screening cell type-specific highly expressed genes

Based on the above data clustering and cell-type identification, we firstly analyzed the expression of cell type-specific genes. The gene-expression proportions in the specific type of cells were calculated. The highly expressed genes were screened according to the following requirements: (1) The gene-expression proportions detected in either of the two populations (sender cells and receiver cells) should be larger than a certain threshold (set to 0.05 by default); (2) The ratio of gene-expression proportions between the two populations should be larger than a certain threshold (set to -Inf by default); (3) In addition, the difference of the mean expression values between the two populations should be larger than a certain threshold (e.g. 0.15). For genes screened by the above criteria, the corresponding expression matrix was then normalized by LogNormalize method. *T* test was used to select highly expressed genes in one versus another cell type (P value < 0.05). These highly expressed genes were considered to be related genes involved in cell-cell interactions. We assumed that the highly expressed genes are most likely to be affected and the most physiologically significant signals in cell interactions, in order to reduce the complexity and false positives of signaling network construction.

scMLnet

A schematic illustration of the scRNA-seq data-based multilayer network method is shown in Figure 1. The multilayer network method provides a new tool for modeling cell-cell

communication and microenvironment-mediated gene expression. To this end, we developed an R package, scMLnet, for constructing the scRNA-seq based multilayer network (<https://github.com/SunXQlab/scMLnet>).

Input and output

Before using scMLnet, scRNA-seq data should be processed and clustered to identify cell types for dissecting cell type-specific gene expressions by employing existing methods or tools (e.g. Seurat [16]). scMLnet requires the following information as input: (1) scRNA-seq expression matrix (a Sparse matrix, where rows represent genes, columns represent cells); (2) clustering results containing two columns: cell's barcode and cluster identities; (3) two cluster identities of receiver cells and sender cells respectively. The output of scMLnet has two forms: (1) tabular information of the constructed multilayer network, containing gene pairs connecting each upstream layer and downstream layer (i.e. ligand-receptor links, receptor-TF links and TF-target links); (2) graphical visualization of the constructed multilayer networks.

Below we describe the algorithmic details of inference and integration of ligand-receptor subnetwork, receptor-TF subnetwork and TF-target gene subnetwork in scMLnet.

Constructing ligand-receptor subnetwork

We collected ligand-receptor pairing information from databases such as DLRP, IUPHAR, HPMR, HPRD, STRING and other databases as well as previous studies [17] to form a list containing 2557

pairs of ligand–receptor directional pairings (Table S1), defined as $ELR = \{(ligandi, receptori)\}$. In order to predict the multilayer signal regulatory network between cell type A (sender cells) and cell type B (receiver cells), we need to obtain the genes that are highly expressed in cell types A and B, respectively (see the ‘Screening cell type-specific highly expressed genes’ section). The ligands with high expression in type A cells were defined as L_H^A , and the receptors with high expression in type B cells were defined as R_H^B , so we select the known ligand–receptor pair for further analysis by searching the pair $(L_H^A \times R_H^B)$ in our ligand–receptor database *ELR*. The selected known ligand–receptor pair $N'LR = ELR \cap (L_H^A \times R_H^B)$ was defined as primary intercellular signaling subnetwork.

Constructing TF–target gene subnetwork

We collected TF–target genes information from TRED, KEGG, TRANSFAC and GeneCards databases. We obtained a TF–target gene list containing 8874 pairs of TF–target gene interactions (Table S1), denoted as $E_{TT} = \{(TF_i, TG_i)\}$. The set of target genes with high expression in type B cells was denoted as TG_{up} , and the list of target genes of a given transcription factor TF_i was denoted as TG_{TF_i} by searching the target genes of a given transcription factor in the TF–target gene list E_{TT} . The collection of all the target genes in the TF–target gene list was denoted as TG_{all} . Fisher’s exact test was used to verify the activation of the transcription factors in cell type B, with the activation probability calculated as:

$$P = \binom{a+b}{a} \binom{c+d}{c} / \binom{a+b+c+d}{a+c} \quad (1)$$

where $\binom{n}{k}$ represents the binomial coefficient. $a = |TG_{up} \cap TG_{TF_i}|$ represents all highly expressed target genes for a given transcription factor, $b = |TG_{up}| - a$ represents the part of the highly expressed genes that are not regulated by this transcription factor, $c = |TG_{TF_i}| - a$ represents the non-highly expressed target genes of TF_i , $d = |TG_{all}| - a - b - c$ represents target genes that are neither highly expressed nor target genes of TF_i . When the P value is lower than 0.05, this transcription factor is considered as a significantly activated transcription factor in receiver cells. The set of transcription factors was defined as TF_A^B . So, the primary subnetwork between transcription factors and target genes can be defined as $N'TT = E_{TT} \cap (TF_A^B \times TG_{up})$.

Constructing receptor–TF subnetwork

The complete signaling network includes ligand–receptor–TF–target gene interactions. In order to connect the primary intercellular subnetwork $N'LR$ with the primary subnetwork between transcription factors and genes $N'TT$, we sought to find the pathways connecting receptors and transcription factors. We used the string database to extract functional links or pathways from the receptors to the downstream transcription factors, including 39 141 pairs of receptor–TF links (Table S1), defined as $ERT = \{(receptor_i, TF_i)\}$. The set of activated transcription factors in type B cells were denoted as TF_A^B , and the collection of transcription factors of a given receptor, R_i , was denoted as TF_{R_i} . The collection of transcription factors was denoted as TF_{all} . Similarly, Fisher’s exact test was used to verify the activation of the receptors, and the activation probability of receptors was calculated by the

following formula:

$$P = \binom{x+y}{x} \binom{m+n}{m} / \binom{x+y+m+n}{x+m} \quad (3)$$

where $\binom{n}{k}$ represents the binomial coefficient. $x = |TF_A^B \cap TF_{R_i}|$ represents all activated transcription factors regulated by a given receptor, $y = |TF_A^B| - x$ is the part of activated transcription factors that not regulated by this receptor, $m = |TF_{R_i}| - x$ represents the non-activated transcription factors regulated by this receptor, $n = |TF_{all}| - x - y - m$ is the transcription factors that are neither activated nor regulated by R_i . When the P value is lower than 0.05, this receptor R_i is considered to be significantly activated. The set of activated receptors was denoted as R_A^B . Therefore, the primary receptor–TF subnetwork can be defined as $N'RT = ERT \cap (R_A^B \times TF_A^B)$.

Refining subnetworks and constructing multilayer signaling network

According to the above analysis, the significantly highly expressed (see ‘Constructing ligand–receptor subnetwork’ section) and activated (see ‘Constructing receptor–TF subnetwork’ section) receptors in cell type B could be defined as $R_{AH} = R_A^B \cap R_H^B$. As such, we could refine ligand–receptor subnetwork and receptor–TF subnetwork to be $N_{LR} = ELR \cap (L_H^A \times R_{AH})$ and $N_{RT} = ERT \cap (R_{AH} \times TF_A^B)$, respectively. More specifically, the ligand–receptor links and receptor–TF links containing receptors that are not highly expressed nor significantly activated were removed from the network. We denoted the corresponding downstream transcription factors in NRT as TF_{NRT} , so the TF–target gene subnetwork could be refined as $N_{TT} = E_{TT} \cap (TF_{NRT} \times TG_{up})$. More specifically, the TF–target gene links containing TFs that were absent in the refined receptor–TF subnetwork NRT were removed.

Furthermore, we performed correlation analysis based on the single cell gene expressions to verify the reliability of the intracellular subnetworks and to further refine the subnetworks. We calculated correlations for each of receptor–TF links and TF–target gene links based on the raw count matrix of the receiver cells using Kendall’s rank–correlation coefficient since it is more robust on non-normal distributions compared with Pearson’s correlation coefficient (PCC). Significantly correlated receptor–TF links (Kendall’s tau $\neq 0$ and P value < 0.05) and TF–target gene links with significant positive correlations (Kendall’s tau > 0 and P value < 0.05) were retained and others were filtered. For simplicity, the refined receptor–TF subnetwork and TF–target gene subnetwork were still denoted as N_{RT} and N_{TT} , respectively.

At last, the multilayer inter-/intra-cellular signaling network could be thereby constructed by integrating the three subnetworks as $N = N_{LR} \cup N_{RT} \cup N_{TT}$. The final ligand–receptor subnetwork N_{LR} and the final receptor–TF subnetwork N_{RT} were connected via the common receptors across the two subnetworks. The final receptor–TF subnetwork N_{RT} and TF–target gene subnetwork N_{TT} were connected via the common TFs. As a result, a directed graph $G = (V, E)$ was constructed, where $V = V(G)$ is the set of vertices consisting of genes in the multilayer network and $E = E(G)$ is the set of edges consisting of ligand–receptor links, receptor–TF links and TF–target gene links.

Applying scMLnet to analyze microenvironment-mediated ACE2 regulation

We applied scMLnet method to analyze microenvironment-mediated ACE2 regulation based on the above BALF scRNA-seq data of COVID-19 patients (GSE145926). We extracted signaling pathways connecting microenvironmental ligands from various types of cells to intracellular ACE2 target gene through receptors and transcription factors. Cells with high expression of ACE2 were viewed as the central cells (the receiver cells), and other cells expressing the ligands were viewed as neighbor cells (the sender cells). We used scMLnet to predict and visualize the multilayer networks, which were then integrated into a multicellular network to fully demonstrate the microenvironmental regulation of ACE2 expression.

Experimental validation

Cells and cytokine treatment

To validate the predicted regulation of cytokines on ACE2 expression in lung cells, a human bronchial cell line (Beas-2B) was cultured for *in vitro* experiments. Beas-2B cells were treated with different doses of EGF (50, 100, 200 ng/ml), IFN- γ (25, 50, 100 ng/ml) or TNF- α (50, 100, 200 ng/ml). In all experiments, cells were incubated with LPS as the control group. All cytokines were obtained from Peprotech (EGF: Animal-Free Recombinant Human EGF, AF-100-15, 100 μ g; TNF- α : Recombinant Human TNF- α , 300-01A, 10 μ g; IFN- γ : Recombinant Human IFN- γ , 300-02, 20 μ g).

Quantitative real-time polymerase chain reaction analysis

Total RNA was isolated from cells using TRIzol (MRC, TR118-500) and cDNA was made using M-MuLV Reverse Transcriptase (Promega, M1705). The mRNA expression level of ACE2 was detected by quantitative real-time polymerase chain reaction (qPCR) using ChamQ SYBR qPCR Master Mix (Vazyme, Q341-03). All kits were used according to the manufacturer's instructions. The specific primers for the PCR were as follows: ACE2: 5' TGCTCAAACAAGCACTCACG 3', Rev: 5' TGTTCATCATGGGGCACAG 3'. All qPCR experiments were run in triplicate.

Results

scMLnet reveals microenvironmental regulation of ACE2 expression using scRNA-seq data of COVID-19 patients

Using scRNA-seq data of BALF samples from COVID patients and healthy controls, 23 916 genes and 66 452 cells were integrated for cell clustering after data pre-processing. After standard procedure for single-cell transcriptome analysis by Seurat [16], a total of 12 types of cells were identified (Figure 2A), including secretory cells, macrophages, ciliated cells, T cells, B cells, plasma cells, plasmacytoid dendritic cells (pDCs), NK cells, mast cells, myeloid dendritic cells (mDCs), Neutrophils and Doublets cells (CD68 + CD3D+), according to cell type marker genes (Figure S1A). ACE2 was mainly expressed in epithelial cells, especially in secretory cells (Figure 2B and C), which is consistent with previous studies that ACE2 is highly expressed in AT2 cells, a type of secretory cell in the lung tissue [18]. In addition, we found that BALFs of COVID-19 patients in severe stage contained higher expression of ACE2 than those in moderate stage (Figure S1B).

We used the scRNA-seq data of COVID19 patient samples to predict multilayer networks of ACE2 regulation during SARS-CoV-2 infection. The annotated output of the clustering analysis for cells in the infected samples was used as input of scMLnet. Since we focus on the regulatory network of ACE2, the secretory cells with higher expression of ACE2 were considered as receiver cells and the other types of cells were considered as sender cells. Multilayer signaling networks produced by scMLnet depicted regulation of ACE2 in secretory cells by other types of cells (Figure S2). The multilayer signaling network consists of four layers: ligand layer, receptor layer, TF layer and target gene layer (i.e. ACE2). Most cell types (including B cells, macrophages, mast cells, mDCs, neutrophils, NK cells, plasma cells and T cells) were predicted to interact with secretory cells, regulating ACE2 expression. Among them, macrophages, T cells and mDCs have more complex intercellular interactions with secretory cells. The intercellular signaling pathways (ligand-receptor interaction) mainly involve the tumor necrosis factor superfamily (TNF, LTA, LTB, TNFSF12), growth factor family (TGFB, AREG), interleukin family (IL1B, IL6) and their respective receptors.

From the above results, we know that scMLnet does not only model intercellular communications but also infer intracellular gene regulatory networks. Moreover, scMLnet used the downstream gene expression to infer activation of the TFs, which was further used to evaluate the activation of the upstream signaling pathways. In this way, only functionally effective ligand-receptor interactions that could transfer the intercellular communications into the intracellular signaling events regulating gene expressions would be retained; other ligand-receptor pairs despite highly expressed would be pruned.

An integrated multilayer network between secretory cells and other types of cells was constructed to investigate the inter-/intra-cellular pathways of ACE2 regulation (Figure 2D). The constructed network demonstrated that multiple signaling pathways, that connected extracellular signals from microenvironmental cells to receptors then to TFs, could regulate ACE2 expression. For instance, AREG that were highly expressed and secreted by Mast cells could bind to the ERBB3 receptor on the surface of secretory cells, and then activate the downstream pathways and TFs, such as c-Fos, STAT1 and HIF1A, regulating ACE2 expression. Likewise, ligand IFNG, highly expressed in NK cells, could bind with receptor IFNGR2 in secretory cells to further activate the downstream pathways and TFs that regulated ACE2 expression. The above analysis demonstrated that ACE2 expression in COVID19 patient BALF was mainly regulated by several specific upstream transcription factors including FOS, HIF1A, STAT1, IRF1. Importantly, the network analysis suggested that these ACE2-regulating transcription factors were activated by some extracellular cytokines, such as AREG and IFNG secreted from immune cells, via the corresponding inter- and intra-cellular signaling pathways.

Verification of ACE2-regulating pathways using bulk RNA-seq data of COVID19 patients

To verify which signaling pathways inferred by the multilayer networks were significantly activated during SARS-CoV-2 infection, we used independent bulk RNA-seq data [19] for analysis. The RNA-seq data for two BALF samples from patients (each in duplicate) were downloaded from the National Genomics Data Center (<https://bigd.big.ac.cn/>) under accession number PRJCA002326. The RNA-seq data for three BALF samples from healthy controls were downloaded from the SRA database under accession numbers: SRR10571724, SRR10571730 and

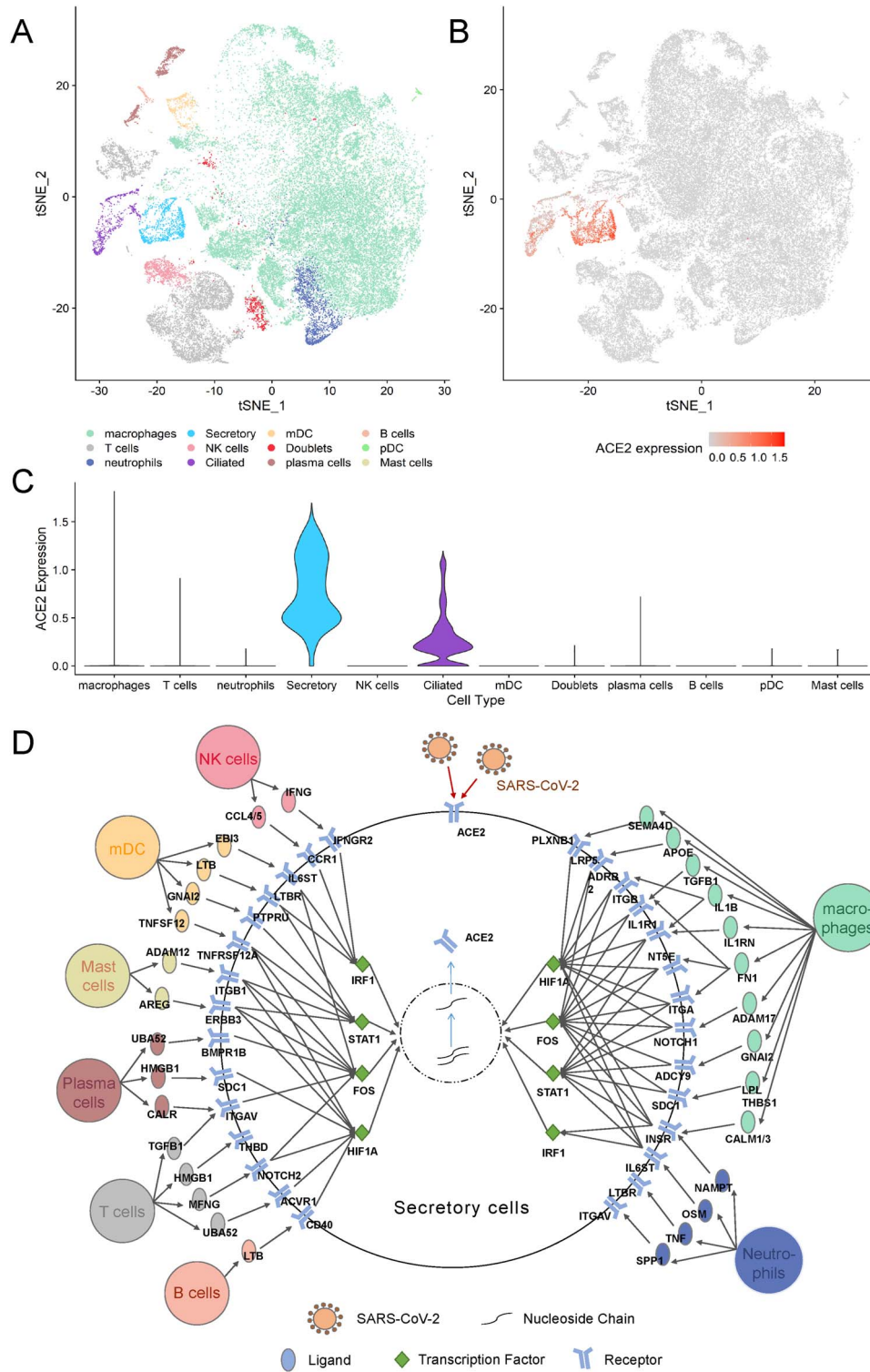


Figure 2. scMLnet applied to scRNA-seq data of BALF in COVID-19 patients reveals microenvironmental regulation of ACE2 expression. (A) Cell-type clustering and identification. According to the expression of marker genes (Figure S1), a total of 12 cell types were obtained, including secretory cells, macrophages, ciliated cells, T cells, B cells, plasma cells, mDC, NK cells, mast cells, pDC cells, neutrophils and doublets cells (CD68 + CD3D+). (B) tSNE plot of ACE2 expression in different cell types. (C) Violin plots of ACE2 expression in different cell types. For visualization purpose, the expression matrix was imputed using 'Run_ALRA' function in Seurat to visualize ACE2 expression. (D) Multicellular network depicting regulation of ACE2 in secretory cells by other types of cells. See related multilayer networks in Figure S2.

SRR10571732. The raw counts were mapped to the human genome (GRCh38) by STAR and then quantitatively analyzed by RSEM with the default parameters. We perform quality

control by filtering out genes that were not expressed in all samples. DESeq2 was used to analyze differential expressions based on the raw counts from transcriptome sequencing data.

Differentially expressed genes (DEGs) were selected by the following criteria: adjusted P -value < 0.01 and fold-change > 4 . Figure 3A shows heatmap of the top 200 significantly up-regulated and down-regulated genes in BALF of COVID-19 patients compared to the healthy controls.

We then used Fisher's exact test (similar to Equation (1)) to assess the statistical significance of activation pathways based on the norm value calculated by DESeq2 and the DEGs (requiring reads counts greater than 10). Pathway information in KEGG database was collected from graphite package v1.30.0. We selected KEGG pathways overlapped with the multilayer networks for analysis. More specifically, the selected pathways should consist of the components that could reach activated transcription factors and that could be reached by the expressed ligands for evaluation. For the Fisher's exact test of each pathway, the significant up-regulated genes and activated TFs were considered up-regulated regardless of their expression level. The ligands and receptors in each pathway were excluded for evaluation since we only focused on the downstream signaling of the pathway. Those pathways with $P < 0.05$ were considered to be significantly activated (Figure S3).

As a result, several pathways were assessed to be significantly activated (Figure 3B), including PI3K-Akt signaling pathway, JAK-STAT signaling pathway, TNF signaling pathway, MAPK signaling pathway and NF- κ B signaling pathway. The overlapped ligand-receptor genes and TFs between the significantly activated pathways and the multilayer networks were also shown in Figure 3B. The significantly activated TFs overlapped with the inferred multilayer network include FOS, HIF1A, STAT1 and IRF1, which were predicted to be major players in regulating ACE2 transcription. The integration of these pathways and TFs suggested the regulatory mechanisms of ACE2 expression induced by extracellular cytokines during SARS-CoV-2 infection (Figure 3C). More specifically, IFN- γ might induce the expression of ACE2 through the downstream transcription factors HIF1A and STAT1 through JAK-STAT signaling pathway and JAK-STAT signaling pathway. AREG (a member of the epidermal growth factor family) could promote expression of ACE2 via the MAPK signaling pathway and JAK-STAT signaling pathway. In addition, TNF- α could also affect ACE2 expression through interacting with specific receptors, LTBR, and the downstream TNF- α signaling pathway and NF- κ B signaling pathway.

Experimental validation of ACE2 regulation by EGF, IFN- γ and TNF- α

Our multilayer network-based predictions of ACE2 regulation are consistent with previous studies. The previous experiments have validated that the expression of ACE2 is not only affected by coronavirus [20, 21] but also regulated by inflammatory cytokines such as IL-1 β [22] and CCL4 [23], which were also predicted by our multilayer network (Figure 2). In addition, the integrated multicellular network further revealed that NK cells might be source cells secreting CCL4 and that macrophages might be source cells secreting IL-1 β (IL1B).

Furthermore, our method made some new predictions on microenvironmental regulation of ACE2 expression during SARS-CoV-2 infection based on scRNA-seq data. As also verified by the bulk RNA-seq data (Figure 3), cytokines EGF, IFN- γ (IFNG) and TNF- α (TNF) were predicted to regulate ACE2 expression. We therefore sought to validate the above predictions using in vitro experiments. The human bronchial epithelial cells (BEAS-2B) were treated with each of the above cytokines at different concentrations. qPCR was performed to measure the expression

levels of ACE2 at 0, 12, 24, 36 and 48 h after cytokine stimulations. The raw data of qPCR were provided in Table S2.

The regulation of ACE2 expression in response to the stimulation of these cytokines in BEAS-2B was confirmed (Figure 4). Stimulation of BEAS-2B with EGF, IFN- γ and TNF- α resulted in time-dependent upregulation of ACE2 expression across all the three concentrations of each cytokine. Among the three cytokines, IFN- γ induced most significant increase of ACE2, particularly at 36 and 48 h. For different concentrations of the three cytokines, the overall trend of ACE2 expression was ascending within 36 h. Interestingly, EGF, IFN- γ and TNF- α induced different temporal patterns of ACE2 expression. High concentrations (100, 200 ng/ml) of EGF resulted in pulse changes of ACE2 expression (i.e. increasing from 0 to 36 h then decreasing afterwards) (Figure 4A). While high concentrations (50, 100 ng/ml) of IFN- γ induced sustained increase of ACE2 expression (Figure 4B). In addition, although TNF- α induced weakest changes of ACE2, stimulation with higher concentration of TNF- α resulted in more elevated ACE2 expression at 48 h (Figure 4C). These experimental data validated the up-regulation of ACE2 by cytokines EGF, IFN- γ and TNF- α , supporting the above multilayer network inference and prediction. These results also provided implications that SARS-CoV-2 could accelerate its entrance into the infected cells by leveraging up-regulation of its receptor gene ACE2 via induction of cytokine storm involving IFN- γ , TNF- α as well as EGF.

Sensitivity of scMLnet to the variation of clustering results

To evaluate the sensitivity or robustness of scMLnet to the variation of the clustering results, we tested the influence of adjusting resolution parameter (set as 0.4, 0.8, 1.2, 1.6, 2) of the 'FindClusters' function in Seurat on network inference. We used the above BALF scRNA-seq data (GSE145926) for evaluation. To identify subtypes of some cell types (e.g. macrophages and T cells) when clustering resolution changed, we used more cell markers: Macrophages (FABP4+ Macro: FABP4, CD52, MARCO; FCN1+ Macro: S100A8, FCN1, S100A9; LGMN+ Macro: SPP1, LGMN, CCL18) and T cells (CCR7+ T: CCR7, IL7R; CD8+ T: GZMA, GZMB; Mixed T: CTLA4, IL2RA) in the workflow for scRNA-seq data analysis (see Methods section). The cell clustering results generated by these different resolution parameters (Figure S4) were used as different inputs of scMLnet.

We focused on the multilayer network prediction of ACE2 regulation, we counted the occurrence times of ligands, receptors, transcription factors and their links in different output networks, so as to examine the sensitivity/robustness of scMLnet to different resolution parameters (res=0.4, 0.8, 1.2, 1.6, 2, respectively). Figure 5A shows the overlaps of ligands, receptors and transcription factors among different multilayer signaling networks. We found that most ligands, receptors and TFs were overlapped among different networks predicted by scMLnet. scMLnet could get the same ligands (52/69), receptors (53/73) and transcription factors (4/4) as ACE2 regulators under three or more different resolutions. Figure 5B shows numbers of predicted ligands, receptors and TFs in the multilayer networks with different resolution parameters. The number of TFs was more robust than that of ligands and receptors with respect to the variations of resolutions. Figure 5C shows the overlaps of ligand-receptor links and receptor-TF links among different multilayer signaling networks with different resolutions. Again, most ligand-receptor links and receptor-TF links were overlapped among different networks predicted by scMLnet. scMLnet could predict the same

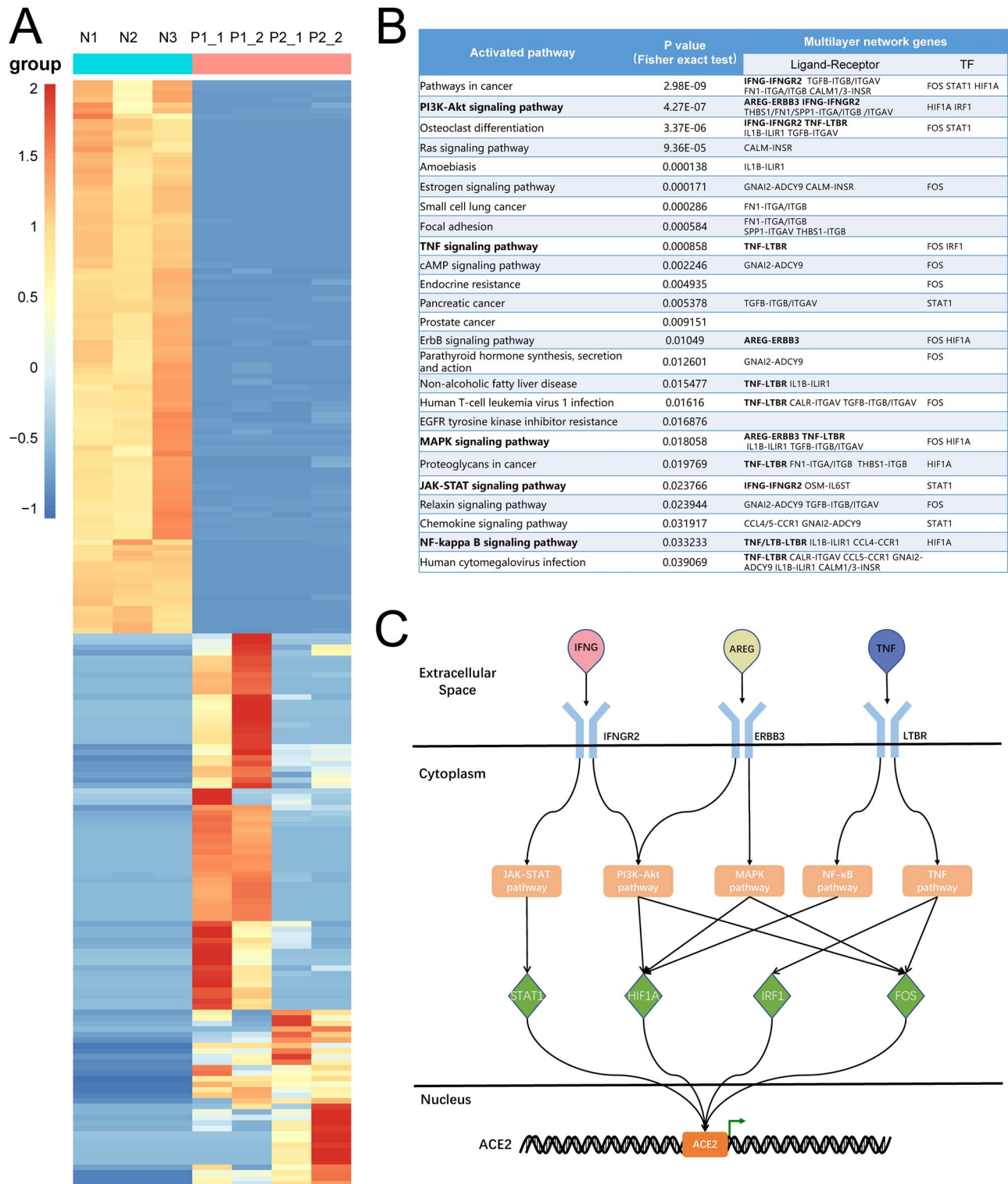


Figure 3. Verification of the predicted ACE2-regulating pathways using independent bulk RNA-seq data of BALF samples from COVID-19 patients. (A) Heatmap of the top 200 significantly up-regulated and down-regulated genes (fold change >2 , $P < 0.01$) in BALF of COVID-19 patients compared to healthy controls. (B) Significantly activated pathways in COVID-19 samples and the overlapped genes in the inferred multilayer networks. The Fisher's exact test was used to assess the significance of activation of each KEGG pathway based on the DEGs. The overlapped ligand-receptor genes and TFs between the significantly activated pathways and the multilayer networks were also listed. (C) A diagram illustrating ACE2-regulating pathways that were significantly activated in COVID-19 patients.

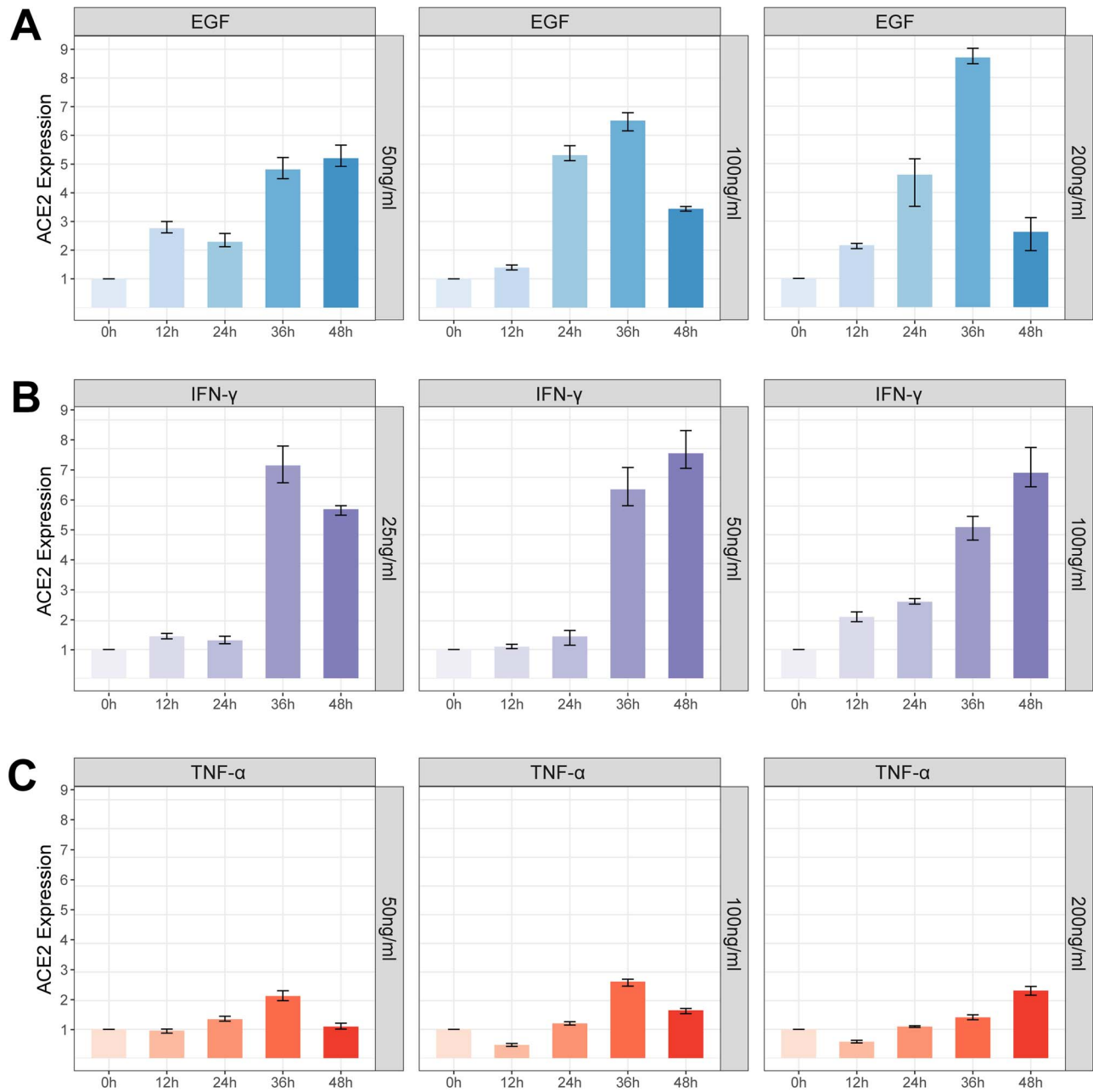


Figure 4. qPCR validation for the regulation of ACE2 expression by cytokines EGF, IFN- γ and TNF- α in human lung cells. The human bronchial epithelial cells (BEAS-2B) were treated with each of the above cytokines for 0, 12, 24, 36 and 48 h. (A) Time-course expression of ACE2 after stimulation with different doses of EGF (50, 100, 200 ng/ml). (B) Time-course expression of ACE2 after stimulation with different doses of IFN- γ (25, 50, 100 ng/ml). (C) Time-course expression of ACE2 after stimulation with different doses of TNF- α (50, 100, 200 ng/ml).

ligand-receptor links (88/139) and receptor-transcription factor links (102/148) under three or more different resolutions.

Discussion

Summary and comparison of different methods

We summarized different methods of inferring cell-cell communication signaling networks in Table 1. Among these methods, SoptSC [8], SingleCellSignalR [9], CellPhoneDB [10], SpaOTsc [11], CellChat [12] and iTALK [24] are mainly designed for inferring intercellular signaling networks, while NicheNet [13], CytoTalk

[25], CCCEXplorer [26] and our method scMLnet are designed to infer inter- and intra-cellular signaling networks. The differences between these methods in input, output, main goal, prior information and main algorithm are also summarized in Table 1.

Based on the case study of ACE2 regulation, we compared scMLnet with other methods (including CCCEXplorer [26] and NicheNet [13]) for inferring extracellular regulations of gene expression. CytoTalk is another method recently developed to infer inter-/intra-cellular signaling network from scRNA-seq data, but we found that it required much more time (over 6 h) to run on a dataset with moderate size (10 000 genes X 100 cells), so we did not include it for comparison. The same scRNA-seq

Table 1. Summary of different methods of inferring cell-cell communication signaling networks.

Methods	Input	Output	Species	Main goal	Prior information	Main algorithm	Ref.	Web site
SoptSC	<p>1. scRNA-seq data: gene-expression matrix (raw) with rows as genes (gene symbols) and columns as cells.</p> <p>2. A vector of gene names.</p> <p>3. A vector of cell ids.</p> <p>4. A vector of cell clusters [optionally].</p> <p>5. Information of pathway of interest: two tables needed, one specifying all unique ligand-receptor combinations and another specifying receptor-target-up/down regulation combinations.</p>	<p>Graph format:</p> <p>1. Chord plot: Ligand-receptor pairs / regulatory direction of the pathway of interest in each of the interacting pairs of cells.</p>	-	Inference of intercellular signaling networks (Ligand-receptor / regulatory direction)	-	<p>Inferring communication networks between single cells using a probability model based on cell-specific expression of ligands, receptors, and target genes in a pathway of interest.</p>	[8]	<p>https://github.com/WangShuxiong/SoptSC</p>
SingleCellSignalR	<p>1. scRNA-seq data: gene-expression matrix (raw) with rows as genes (gene symbols) and columns as cells.</p> <p>2. Cell type annotation: a vector of cell clusters (from other tools or SingleCellSignalR)</p>	<p>Tabulate format:</p> <p>1. Ligand-receptor information: the first column is ligand, the second is receptor, and the others are attributes of genes/interactions.</p> <p>Graph format:</p> <p>1. Chord plot: number/score of Ligand-receptor pairs in each of the interacting pairs of cells.</p> <p>2. Intercellular network including ligands and receptors.</p>	Human or Mouse	Inference of intercellular signaling networks (Ligand-receptor)	3251 ligand-receptor pairs by integrating and curating existing sources plus manual additions	<p>A regularized score to assess the confidence in predicted ligand-receptor interactions by controlling false positives.</p>	[9]	<p>https://github.com/SCA-IRCM</p>

Continued

Table 1. Continued

Methods	Input	Output	Species	Main goal	Prior information	Main algorithm	Ref.	Web site
CellPhoneDB	<p>1. scRNA-seq data:gene-expression matrix (recommended normalized) with rows as genes (ensemble IDs) and columns as cells.</p> <p>2. Cell type annotation:two columns (the first column is barcode and the second is cell type).</p>	<p>Tabulate format: 1. Ligand-receptor information:the first two columns are ligand-receptor pairs information, the third to the eighth columns are ligand or receptor attributes, and the others are the p.val referring to the enrichment of the ligand-receptor pairs in each of the interacting pairs of cells.</p> <p>Graph format: 1. Heatmap: number of Ligand-receptor pairs in each of the interacting pairs of cells.</p> <p>2. Bubble plot: significance of ligand-receptor pairs in each of the interacting pairs of cells.</p>	Human or Mouse	Inference of intercellular signaling networks (ligand-receptor)	Ligand-receptor information from databases and previous studies: 1. Secreted and membrane proteins: 501 secreted proteins and 585 membrane proteins. 2. Protein complexes: 112 pairs. 3. Protein-protein interactions: 1396 are interactions and 466 are heteromers.	Predicting enriched ligand-receptor interactions considering the subunit architecture of both ligands and receptors by empirical shuffling and statistical test.	[10]	https://github.com/Teichlab/cellphonedb
SpaOTsc	<p>1. scRNA-seq data:gene-expression matrix (raw) with rows as cells (gene symbols) and columns as genes.</p> <p>2. Spatial data:distance matrix among spatial locations.</p> <p>3. Dissimilarity matrix between scRNA-seq data and spatial data.</p> <p>4. Dissimilarity matrix within scRNA-seq data.</p> <p>5. Information of pathway of interest: ligands, receptors and target genes.</p>	<p>Tabulate format: 1. Probability matrix indicating cell i sending signal to cell j in pathway of interest.</p> <p>Graph format: 1. Heatmap plot: pathway probability summarized into cell clusters. 2. UMAP plot:regulatory direction of the pathway of interest in each of the interacting pairs of cells/clusters.</p>	-	Inference of intercellular signaling networks (regulatory direction)	-	Using structured and unbalanced optimal transport to recover spatial properties of scRNA-seq data in the aid of the spatial reference and then 'optimally' transporting' signal senders to target signal receivers.	[11]	https://github.com/zcang/SpaOTsc

Continued

Table 1. Continued

Methods	Input	Output	Species	Main goal	Prior information	Main algorithm	Ref.	Web site
CellChat	<p>1. scRNA-seq data: gene-expression matrix (normalized) with rows as genes (gene symbols) and columns as cells.</p> <p>2. Cell type annotation (label-based mode) or low-dimensional representation of scRNA-seq data such as principal component analysis or diffusion map (label-free mode).</p>	<p>Graph format:</p> <ol style="list-style-type: none"> Hierarchical plot: the inferred intercellular communication network. Circle plot: number of ligand-receptor pairs in each of the interacting pairs of cells. Bubble plot: significance of ligand-receptor pairs in each of the interacting pairs of cells. 	Human or Mouse	Inference of intercellular signaling networks (Ligand-receptor)	Ligand-receptor information: 2021 validated molecular interactions including complexes with multimeric ligands and receptors, as well as several cofactors types: soluble agonists, antagonists, co-stimulatory and co-inhibitory membrane-bound receptors	Quantitatively inference of intercellular communication networks from scRNA-Seq data by the law of mass action and statistical test.	[12]	https://github.com/sqjin/CellChat
iTALK	scRNA-seq data: gene-expression matrix (raw/normalized) with rows as cells and columns as genes (gene symbols). One of the matrix columns is cell type annotation.	<p>Tabulate format:</p> <ol style="list-style-type: none"> Ligand-receptor information: the first column is ligand, the second is receptor, and the others are gene attributes. <p>Graph format:</p> <ol style="list-style-type: none"> Network plot: the number of ligand-receptor interactions in each of the interacting pairs of cells. Circle plot: ligand-receptor interactions in each of the interacting pairs of cells. <p>2. Errorbar plot: the dynamic changes of a certain ligand-receptor pair across multiple timepoints or subgroups.</p>	Human	Inference of intercellular signaling networks (ligand-receptor)	Ligand-receptor information: 2648 non-redundant and known interacting ligand-receptor pairs (from previous studies) which were further classified into four categories based on the primary function of ligand: cytokines/Chemokines, immune checkpoint genes, growth factors, and others.	Identifying and illustrating (dynamic) intercellular communication according to the study design (at a single timepoint/group or at multiple timepoints/genetically subgroups) by gene expression.	[24]	https://github.com/Coolgenome/iTALK

Continued

Table 1. Continued

Methods	Input	Output	Species	Main goal	Prior information	Main algorithm	Ref.	Web site
NicheNet	1. Bulk or single cell RNA-seq data:gene-expression matrix (normalized)with rows as genes (gene symbols) and columns as cells. 2. Cell type annotation:meta-information about the cells including barcodes and cell types. 3. Cell type:senders (A) and receivers (B). 4. Target gene set of interest.	Tabulate format: 1. Signaling subnetworks:the first column is source node, the second is target node, and the third is edge weight. 2. Regulatory subnetworks:the first column is TF, the second is target gene, and the third is edge weight. Graph format: pathway information including signaling mediators and TFs between given ligands and target genes.	Human or Mouse	Inference of inter- and intra-cellular signaling networks (Ligand-Signaling mediator, TF-Target gene)	Prior model of ligand-target regulatory potential model built by integrating prior knowledge on ligand-to-target signaling paths (from 57 child databases)	Inferring signaling subnetworks (Ligand-Signaling mediator) and regulatory subnetworks (TF-Target gene) based on gene expression and Ligand-Target regulatory potential model	[13]	https://github.com/saeyslab/nichenetr
CytoTalk	1. scRNA-seq data:gene-expression matrix (log2-transformed)of specific cell types with rows as genes (gene symbols) and columns as cells. 2. Species information:Human or Mouse	Tabulate format: 1. Edge attribute:edge cell type, edge cost. 2. Node attribute:gene cell type, gene expression, gene prize, gene real name. graph format: Crosstalk network including ligands, receptors and signaling mediators.	-	Inference of inter- and intracellular signaling network (Ligand-receptor-Signaling mediator)	-	Using an information-theoretic measure between two cell types to construct intra- and intercellular gene-gene networks; and using the prize-collecting Steiner forest algorithm to identify candidate signal pathways.	[25]	https://github.com/tanlabcode/CytoTalk

Continued

Table 1. Continued

Methods	Input	Output	Species	Main goal	Prior information	Main algorithm	Ref.	Web site
CCCExplorer	Bulk RNA-seq data: gene-expression matrix (FPKM) of specific cell types (sender cells (A) and receiver cells (B)) with rows as genes (gene symbols) and columns as cells. Matrix includes two conditions (control and experimental) and one of the columns is qval obtained by differential expression analysis between the control conditions and the experimental conditions.	Tabular format: 1. Edge attribute: the first column is source node, the second is target node, and the others are edge attributes 2. Node attribute: the first column is node name, the second is node type (ligand/receptor/TF/signaling mediator), and the third is pathway. Image format: Crosstalk network including ligands, receptors, signaling mediators and TFs.	Human	Inference of inter- and intra-cellular signaling network (ligand-receptor-signaling mediator-TF)	1. Ligand-receptor information: 1433 pairs (from DLRP, KEGG, HPRD, STRING databases and previous studies). 2. Receptor-TF information: KEGG databases. 3. TF-Target gene information: KEGG, TRED databases.	Integrating inter-/intra-cellular pathways based on prior network information, statistical inference (Fisher's exact test) and a directed graph.	[26]	https://github.com/methodistsmab/CCCExplorer
scMLnet	1. scRNA-seq data: gene-expression matrix (raw) with rows as genes (gene symbols) and columns as cells. 2. Cell type annotation: two columns. The first column is barcode and the second is cell type. 3. Cell type: sender cells (A) and receiver cells (B).	Tabulate format: 1. Ligand-receptor subnetworks: a column, connecting ligands with receptors by underlined for visualization of Ligand-receptor subnetworks. 2. Receptor-TF subnetworks: a column, connecting receptors with TFs by underlined for visualization of Receptor-TF subnetworks. 3. TF-target gene subnetworks: a column, connecting TFs with target genes by underlined for visualization of TF-target gene subnetworks. Graph format: Multi-layer network visualization of ligand-receptor-TF-target gene links.	Human	Inference of inter- and intracellular signaling network (ligand-receptor-TF-target gene)	1. Ligand-receptor information: 2557 pairs (from DLRP, IUPHAR, HPMR, HPRD, STRING databases and previous studies) 2. Receptor-TF information: 39141 pairs (from STRING databases) 3. TF-target gene information: 8874 pairs (from TRED, KEGG, GeneCards and TRANSFAC databases).	Inferring intercellular subnetworks (Ligand-receptor interactions) and intracellular subnetworks (Receptor-TF interactions and TF-Target gene interactions) based on gene expression prior information and statistical inference (Fisher's exact test and correlation); and integrating subnetworks based on overlapping molecules.	This study	https://github.com/SunXQlab/scMLnet

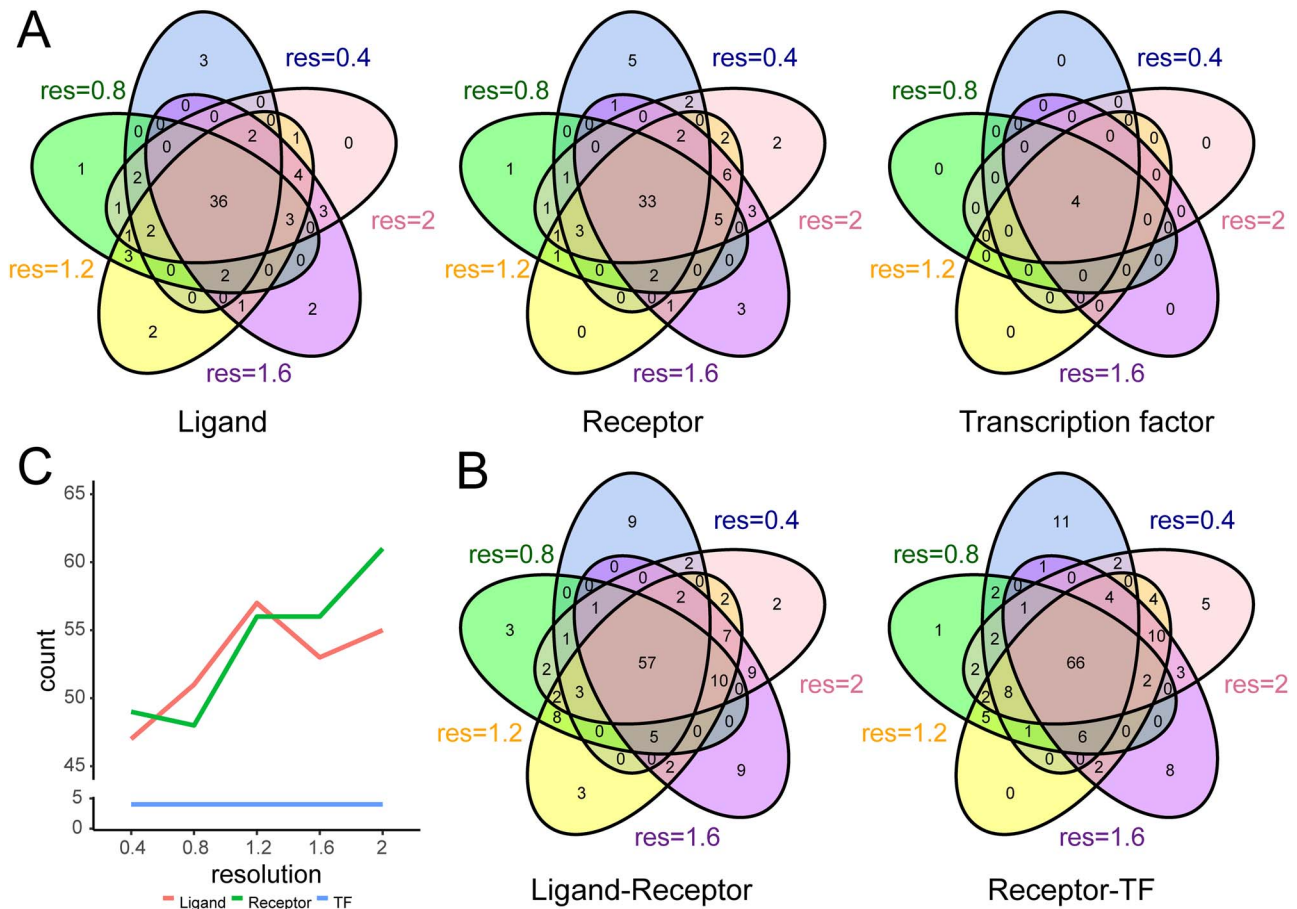


Figure 5. Robustness of scMLnet to cell clustering with different resolution parameters. (A) Overlap of ligands, receptors and transcription factors in different multilayer signaling networks inferred with input from clustering using Louvain algorithm at different resolutions (res=0.4, 0.8, 1.2, 1.6, 2, respectively). (B) Number of ligands, receptors and TFs in the multilayer network regulating ACE2 predicted by scMLnet with different resolutions. (C) Overlap of ligand-receptor links and receptor-TF links in the multilayer signaling network with different resolutions.

dataset of the BALF (GSE145926) was used to test whether the above methods could correctly predict the regulations of ACE2 by EGF, IFN- γ and TNF- α . scMLnet correctly inferred the three ligands regulating ACE2 expression as experimentally validated above, whereas CCExplorer and NicheNet did not (Figure S5).

Here, we focus on discussing the major differences between scMLnet and NicheNet for inferring inter-/intra-cellular signaling network from scRNA-seq data. (1) Difference in the required data for input: NicheNet depends on (bulk or single-cell) gene-expression data with differential expression information, which cannot be ‘steady state’ data (see also in Supplementary Notes Table 1 of [13]). Both target gene set of interest and possible active ligands are needed for NicheNet to perform ligand activity prediction. In contrast, scMLnet uses single-cell gene-expression data as input to infer interactions between any two types of annotated cells, which can be ‘steady state’ data. Moreover, scMLnet does not require predefined target gene set. (2) Difference in algorithm of network inference: NicheNet employs network propagation methods on the integrated networks to calculate a ligand-target regulatory potential score, which depends more on the integrated network of prior information but less on gene-expression values. In contrast, our scMLnet exploits more about the information of single cell gene expressions by employing statistical inference (Fisher’s exact test and correlation test) on the integrated networks to evaluate the significance

of TF activation and pathway activation and to evaluate the correlations of receptor-TF targets links.

Furthermore, we calculated correlations in R-TF target gene links predicted by scMLnet and NicheNet for a quantitative comparison. To this end, we applied scMLnet and NicheNet to the above scRNA-seq dataset of the BALF (GSE145926) to predict upstream regulators and pathways of a COVID-19-related gene set (500 genes up-regulated in A549 cells treated with SARS-CoV-2 at both 0.2 MOI and 2MOI concentrations) that was downloaded from the COVID-19 Gene and Drug Set Library [27] (GSE147507 dataset). This gene set was used as the target gene set of interest for input of NicheNet. For scMLnet (with default parameters), the intersection between the inferred target gene set and the above gene set was used as the final target genes in the multilayer signaling network. We then calculated PCC between genes in each receptor-TF link and TF-target gene link of the two inferred networks (Figure S6). We found that both R-TF correlations and TF-target gene correlations were higher for scMLnet network than NicheNet network, supporting that scMLnet could predict more reliable links.

General applicability of scMLnet

Our method can be applied to at least the following two aspects. First, scMLnet can be used to predict microenvironmental

regulators of gene expression within a cell using scRNA-seq data. It can be used to infer signaling mechanisms underlying microenvironment-mediated regulation of important genes in tumors or developmental tissues. For example, programmed cell death 1 (PD-1) and its ligand programmed death ligand-1 (PD-L1/B7-H1/CD274) play critical roles in T cell-mediated tumor immunity and their expressions have been shown to be regulated by tumor microenvironment [28]. In the future study, we will apply scMLnet to infer which cells and which cytokines in the tumor microenvironment can regulate PD-L1 expression, which may provide insights on designing novel combination immunotherapies. Second, scMLnet can be used to investigate functional roles of intercellular communications. It can be used to investigate how interacting cells via ligand-receptor signaling affect downstream gene expressions that could be enriched in biological process or functions. As such, the intercellular communication networks predicted by scMLnet could be manipulated to control cell fate decisions or may be used as biomarkers to predict disease progression [7].

Limitations of our method

Our method has several limitations which could be improved in the future studies. First, the prior information of ligand-receptor interactions, signaling pathways and gene regulations used in the current study may not be the most comprehensive one. In the future work, we will integrate more available data sources to build a weighted network consisting of ligand-receptor-TF target links using a similar approach of NicheNet. Based on the integrated weighted network, we will further train a predictive model using single-cell gene-expression data. Second, although scRNA-seq data involves a large number of cell samples for statistical analysis and model training, gene-expression profiles from single cells are sparse due to technical artifacts such as dropout, which may lead to bias of quantitative analysis. In the future study, we will incorporate data imputation procedure into our tool to mitigate those impacts.

Despite the above limitations, our multilayer network approach could serve as a hypothesis-generating tool for further experimental validation. The findings in our study are helpful for understanding the mechanism underlying tissue microenvironment-mediated regulation of ACE2 expression. We validated the prediction of ACE2 regulation by cytokines EGF, IFN- γ and TNF- α (Figure 4). In addition, we verified the correlations between expressions of ACE2 and receptor genes and the correlations between transcription factor and receptor genes in ACE2+ cells revealed by the multilayer networks. Note that the expression levels of the ligand genes and ACE2 are not in pairing relationship and hence it is not feasible to calculate their correlations. We calculated the Kendall correlations between 27 receptors, 4 transcription factors and target gene ACE2 in multilayer signal networks based on the gene expressions in ACE2+ cells (Figure S7). These results supported that almost all the predicted receptor TF links and TF-ACE2 regulations were significantly correlated.

Implications for understanding and intervention of COVID-19

The COVID-19 pandemic has posed great threat to human health and caused enormous impact on economic and social development. It is in urgent need to understand the pathogenesis of SARS-CoV-2. ACE2 has been identified as a key receptor of SARS-CoV-2 [29], the same for SARS-CoV [30]. ACE2 is not only a

channel for SARS-CoV-2 to invade cells, but also plays important roles in the development of coronavirus disease [31]. Previous studies have suggested that changes in ACE2 expression has important roles in COVID-19 progression and may be related to inflammatory response to SARS-CoV-2 infection. A recent study showed that type I interferon IFN can elevate ACE2 expression in airway epithelial cells both *in vitro* and *in vivo* [32]. Therefore, unraveling the microenvironmental mechanism of ACE2 regulation is of therapeutic relevance to COVID-19.

Modeling, inference and analysis of molecular regulatory networks is one of the central questions in computational systems biology, which has wide applications in biomedical research [33–35]. For example, network-based approaches have been developed to predict subtype-specific drug targets [36] or to predict tumor recurrence in breast cancers [37] using network propagation (or heat diffusion) methods. In this study, to elucidate how immune microenvironment regulates ACE2 expression, we developed a single-cell RNA-seq transcriptomic data-based multilayer network approach to investigate interactions between ACE2+ cells and microenvironmental cells.

Our study has implications for understanding and intervention of COVID-19. Several studies have shown that patients in intensive care might develop acute respiratory distress syndrome (ARDS), a devastating clinical syndrome with a high mortality rate [38, 39]. With the development of the disease, the over-activated immune system releases a large number of cytokines to induce ARDS [40]. ACE2 and ACE are two important components in renin angiotensin system and have distinct roles in the development of ARDS [31]. Therefore, the dysregulated ACE2 expression mediated by inflammatory immune microenvironment might act as a mechanism of coronavirus disease development and may explain how inflammatory cytokine storm promotes the induction of ARDS in patients with severe COVID-19 who have elevated cytokine levels [41].

Therefore, it is reasonable to propose that the strategy modulating ACE2 expression may serve as effective intervention of COVID-19. As revealed in our study, the ACE2 expression could be controlled by targeting above-mentioned TFs such as c-Fos and STAT1 or manipulating cytokines such as IFN- γ and TNF- α , for controlling COVID-19 severity. In fact, TNF-blocking antibodies, such as adalimumab, etanercept and golimumab, have been reported to effectively treat inflammatory diseases [42]. Clinical trials of anti-TNF- α therapy are thus urgently recommended for COVID-19. Another therapy based on targeting IFN- γ treatment is also promising and feasible by inhibiting the associated JAK-STAT signaling pathways which has been under evaluation in a clinical trial for COVID-19 [43]. As such, therapies targeting c-Fos and STAT1 are strongly recommend and may represent a novel treatment for controlling COVID-19, which is in urgent need of clinical trials.

Conclusion

In conclusion, we developed a scRNA-seq based multilayer network approach to reveal microenvironment-mediated regulation of ACE2 expression for better understanding SARS-CoV-2 entry and cell infection. We revealed and validated that ACE2 expression could be up-regulated by extracellular cytokines EGF, IFN- γ and TNF- α , via PI3K-Akt pathway, JAK-STAT pathway and MAPK pathway as well as their downstream transcription factors. Our study brings new insights into tissue microenvironment-mediated inter-/intra-cellular signaling mechanisms of ACE2 regulation and provides implications for designing rational therapeutic strategies for COVID-19.

We summarized and compared different methods of inferring signaling networks underlying cell–cell communications from scRNA-seq data, providing a valuable roadmap for tools' applications and new methodology development.

Key Points

- The scRNA-seq data based multilayer network (scMLnet) method provides a new tool for modeling both intercellular communications and intracellular gene regulatory networks, with an R package available at <https://github.com/SunXQlab/scMLnet>.
- scMLnet could be used to predict microenvironmental regulators (e.g. cell types and cytokines) of gene expression using scRNA-seq data and to understand functional roles of intercellular communications.
- Applying scMLnet to a COVID-19 dataset, we predicted that ACE2 expression could be up-regulated by extracellular cytokines EGF, IFN- γ and TNF- α , which was validated using in vitro experiments in human lung cells.
- Our study brings new insights into immune microenvironment-mediated regulation of ACE2 expression and provides implications for designing novel therapeutic strategies of COVID-19.

Data availability

The data used for analysis in this study were downloaded from publicly available databases that have been described in the main text. The R code used for analysis and the scMLnet package are available at <https://github.com/SunXQlab/scMLnet>.

Funding

X.S. was supported by grants from the National Natural Science Foundation of China (11871070), the Guangdong Basic and Applied Basic Research Foundation (2020B1515020047) and the Fundamental Research Funds for the Central Universities (20ykzd20).

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Boulanger CA, Mack DL, Booth BW, et al. Interaction with the mammary microenvironment redirects spermatogenic cell fate in vivo. *Proc Natl Acad Sci U S A* 2007;104:3871–6.
2. Quail DF, Joyce JA. The microenvironmental landscape of brain tumors. *Cancer Cell* 2017;31:326–41.
3. Sun X, Su J, Bao J, et al. Cytokine combination therapy prediction for bone remodeling in tissue engineering based on the intracellular signaling pathway. *Biomaterials* 2012;33:8265–76.
4. Livesey FJ, Cepko CL. Vertebrate neural cell-fate determination: lessons from the retina. *Nat Rev Neurosci* 2001;2:109–18.
5. Ammeux N, Housden BE, Georgiadis A, et al. Mapping signaling pathway cross-talk in *Drosophila* cells. *Proc Natl Acad Sci U S A* 2016;113:9940.
6. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol* 2018;18:35–45.
7. Zhang J, Guan M, Wang Q, et al. Single-cell transcriptome-based multilayer network biomarker for predicting prognosis and therapeutic response of gliomas. *Brief Bioinform* 2020;21:1080–1097.
8. Wang S, Karikomi M, MacLean AL, et al. Cell lineage and communication network inference via optimization for single-cell transcriptomics. *Nucleic Acids Res* 2019;47:e66–6.
9. Cabello-Aguilar S, Alame M, Kon-Sun-Tack F, et al. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res* 2020;48:e55–5.
10. Efremova M, Vento-Tormo M, Teichmann SA, et al. Cell-PhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat Protoc* 2020;15:1484–506.
11. Cang Z, Nie Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat Commun* 2020;11:2084.
12. Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell–cell communication using CellChat. 2020; *BioRxiv* 2020.07.21.214387.
13. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods* 2020;17:159–62.
14. Liao M, Liu Y, Yuan J, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med* 2020;26:842–4.
15. Morse C, Tabib T, Sembrat J, et al. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur Respir J* 2019;54(2):1802441.
16. Satija R, Farrell JA, Gennert D, et al. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 2015;33:495–502.
17. Choi H, Sheng J, Gao D, et al. Transcriptome analysis of individual stromal cell populations identifies stroma-tumor crosstalk in mouse lung cancer model. *Cell Rep* 2015;10:1187–201.
18. Zhao Y, Zhao Z, Wang Y, et al. Single-cell RNA expression profiling of ACE2, the receptor of SARS-CoV-2. *Am J Respir Crit Care Med* 2020;202:756–9.
19. Xiong Y, Liu Y, Cao L, et al. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerg Microbes Infect* 2020;367:1260–1263.
20. Li Y, Li H, Zhou L. EZH2-mediated H3K27me3 inhibits ACE2 expression. *Biochem Biophys Res Commun* 2020;526:947–952.
21. Kuba K, Imai Y, Rao S, et al. A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat Med* 2005;11:875–9.
22. Clarke NE, Belyaev ND, Lambert DW, et al. Epigenetic regulation of angiotensin-converting enzyme 2 (ACE2) by SIRT1 under conditions of cell energy stress. *Clin Sci (Lond)* 2014;126:507–16.
23. Zhang W, Miao J, Li P, et al. Up-regulation of components of the renin–angiotensin system in liver fibrosis in the rat induced by CCL4. *Res Vet Sci* 2013;95:54–8.

24. Wang Y, Wang R, Zhang S, et al. iTALK: an R Package to characterize and illustrate intercellular communication. *BioRxiv* 2019;507871.
25. Hu Y, Peng T, Gao L, et al. CytoTalk: De novo construction of signal transduction networks using single-cell RNA-Seq data. *Biorxiv* 2020.03.29.014464.
26. Choi H, Sheng J, Gao D, et al. Transcriptome analysis of individual stromal cell populations identifies stroma-tumor crosstalk in mouse lung cancer model. *Cell Rep* 2015;10:1187–201.
27. Kuleshov MV, Clarke DJB, Kropiwnicki E, et al. The COVID-19 drug and gene set library. *Patterns* 2020;1(6):100090.
28. Wang Y, Wang H, Yao H, et al. Regulation of PD-L1: emerging routes for targeting tumor immune evasion. *Front Pharmacol* 2018;9:536.
29. Wrapp D, Wang H, Yao H, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 2020;eabb2507.
30. Song W, Gui M, Wang X, et al. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog* 2018;14:e1007236.
31. Imai Y, Kuba K, Rao S, et al. Angiotensin-converting enzyme 2 protects from severe acute lung failure. *Nature* 2005;436:112–6.
32. Ziegler CGK, Allon SJ, Nyquist SK, et al. SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. *Cell* 2020;181:1016–1035.
33. Sun X, Bao J, Nelson KC, et al. Systems modeling of anti-apoptotic pathways in prostate cancer: psychological stress triggers a synergism pattern switch in drug combination therapy. *PLoS Comput Biol* 2013;9:e1003358.
34. S Li P, Yang M, Zhang Y, et al. Dissecting the single-cell transcriptome network underlying gastric premalignant lesions and early gastric cancer. *Cell Rep* 2019;27:1934–1947.e5.
35. Ji Z, Guan M, Wang Q, et al. Single-cell transcriptome-based multilayer network biomarker for predicting prognosis and therapeutic response of gliomas. *Brief Bioinform* 2020;21:1080–1097.
36. Zaman N, Li L, Jaramillo ML, et al. Signaling network assessment of mutations and copy number variations predict breast cancer subtype-specific drug targets. *Cell Rep* 2013;5:216–23.
37. Milanese JS, Tibiche C, Zou J, et al. Germline variants associated with leukocyte genes predict tumor recurrence in breast cancer patients. *NPJ Precis Oncol* 2019;3:28.
38. Xu Z, Shi L, Wang Y, et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir Med* 2020;8:420–422.
39. Hudson LD, Milberg JA, Anardi D, et al. Clinical risks for development of the acute respiratory distress syndrome. *Am J Respir Crit Care Med* 1995;151:293.
40. Lan C-C, Huang H-K, Wu C-P, et al. Recruitment maneuver leads to increased expression of pro-inflammatory cytokines in acute respiratory distress syndrome. *Respir Physiol Neurobiol* 2020;271:103284.
41. Long Q-X, Tang X-J, Shi Q-L, et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat Med* 2020;26:1200–1204.
42. Feldmann M, Maini RN, Woody JN, et al. Trials of anti-tumour necrosis factor therapy for COVID-19 are urgently needed. *The Lancet* 2020;395:1407–1409.
43. Stebbing J, Phelan A, Griffin I, et al. COVID-19: combining antiviral and anti-inflammatory treatments. *Lancet Infect Dis* 2020;20:400–2.