

# scDCA: deciphering the dominant cell communication assembly of downstream functional events from single-cell RNA-seq data

Boya Ji<sup>1</sup>, Xiaoli Wang<sup>1</sup>, Xiang Wang<sup>2</sup>, Liwen Xu<sup>1,\*</sup>, Shaoliang Peng<sup>1,\*</sup>

<sup>1</sup>College of Computer Science and Electronic Engineering, Hunan University, Yuelu, 410006 Changsha, China

<sup>2</sup>The Second Xiangya Hospital, Central South University, Yuelu, 410006 Changsha, China

\*Corresponding authors. Liwen Xu, College of Computer Science and Electronic Engineering, Hunan University, Yuelu, 410006 Changsha, China.

E-mail: xuliwen@hnu.edu.cn; Shaoliang Peng, College of Computer Science and Electronic Engineering, Hunan University, Yuelu, 410006 Changsha, China.

E-mail: slpeng@hnu.edu.cn

## Abstract

Cell-cell communications (CCCs) involve signaling from multiple sender cells that collectively impact downstream functional processes in receiver cells. Currently, computational methods are lacking for quantifying the contribution of pairwise combinations of cell types to specific functional processes in receiver cells (e.g. target gene expression or cell states). This limitation has impeded understanding the underlying mechanisms of cancer progression and identifying potential therapeutic targets. Here, we proposed a deep learning-based method, scDCA, to decipher the dominant cell communication assembly (DCA) that have a higher impact on a particular functional event in receiver cells from single-cell RNA-seq data. Specifically, scDCA employed a multi-view graph convolution network to reconstruct the CCCs landscape at single-cell resolution, and then identified DCA by interpreting the model with the attention mechanism. Taking the samples from advanced renal cell carcinoma as a case study, the scDCA was successfully applied and validated in revealing the DCA affecting the crucial gene expression in immune cells. The scDCA was also applied and validated in revealing the DCA responsible for the variation of 14 typical functional states of malignant cells. Furthermore, the scDCA was applied and validated to explore the alteration of CCCs under clinical intervention by comparing the DCA for certain cytotoxic factors between patients with and without immunotherapy. In summary, scDCA provides a valuable and practical tool for deciphering the cell type combinations with the most dominant impact on a specific functional process of receiver cells, which is of great significance for precise cancer treatment. Our data and code are free available at a public GitHub repository: <https://github.com/pengsl-lab/scDCA.git>.

**Keywords:** dominant cell communication assembly; graph convolution network; deep learning; model interpretability; cell-cell communication; single-cell RNA-seq

## Introduction

Cell-cell communications (CCCs) underpin the major functions in multicellular organisms [1]. Sender cells induce cascading regulation of biological processes through CCCs, resulting in the reprogramming of the receiver cell with altered gene expression and functional pathways [2]. Depicting the landscape of CCCs not only provides insights into fundamental life processes but also offers strategies for therapeutic interventions for various complex diseases, especially cancer [3]. Numerous studies have contributed to identifying the communicating cell component pairs that act as the functional assembly inducing downstream biological processes and clinical response. For instance, inter-cellular communication between immune and cancer cells can modulate the expression of immune-related genes and promote tumor evasion [4]. The communication between cancer cells and cancer-associated fibroblasts can induce the expression of matrix metalloproteinases, which are essential for extracellular matrix remodeling and tumor invasion [5].

Recent advances in single-cell RNA sequencing (scRNA-seq) enabled the characterization and interpretation of the complete

landscape of CCC using computational methods. Generally, these computational methods inferred CCC through integrating scRNA-seq data and prior ligand-receptor (L-R) interaction information, and employ various models to assess the enrichment or over-presentation of L-R interactions based on the constructed background reference. Two of the most widely used methods, CellPhoneDB [6] and CellChat [7], applied statistical tests to quantify the probability of each interaction over null hypothesis references. Besides, there are other methods that introduce the information of gene networks for a more accurate and complete model. For instance, NicheNet integrates the intracellular gene regulatory network [8], CytoTalk incorporated intracellular signaling networks between two cell types [9], and NATMI constructed cell-connectivity-summary networks [10]. In addition, spatial information of cells was also incorporated in some methods to refine the cell-cell interactions prediction, such as stLearn [11] and CellPhoneDB v3.0 [12].

However, to the best of our knowledge, current CCC analysis methods lack the ability to identify the communicating cell type pairs that play a major role in a specific downstream functional

Received: October 21, 2024. Revised: November 24, 2024. Accepted: December 4, 2024

© The Author(s) 2024. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

event. A complete characterization of CCC in microenvironment required not only the identification of communicating cell types with the enriched L-R interactions, but also their importance in affecting specific downstream functional events, e.g. the expression pattern of target genes or the functional state of receiver cells. Information exchange between different cells in microenvironment, not all cells have a direct or equal impact on downstream functional events. How to rank the importance of these CCCs and find the dominant ones has great significance for deciphering various physiological and pathogenic processes [1]. Additionally, most of existing methods perform CCC analysis at the level of the cell type or cluster, discarding single-cell-level information. Some communication networks could be obscured with cell aggregation or downsampling [13]. In summary, it is vital to decipher CCC importance in affecting specific downstream functional events at single-cell resolution to accurately gain the full landscape of intercellular communication.

Inspired by previous researches, we noticed that the graph neural networks (GNNs) and the attention mechanism in deep learning are suitable for addressing these challenges. First, GNNs can well model the CCC networks. For example, Fischer et al. [14] utilized GNNs to build node-centric expression modeling (NCEM) for improving CCC inference. Yuan et al. [15] introduced the GCNG method, which utilized the graph convolutional neural network to infer L-R interactions. In addition, the attention mechanism in deep learning can help assess the contribution of different variables in the model and further prioritize the more important ones [16]. For example, Chen et al. [17] developed the GENELink method, which utilized the graph attention network to infer the high-confidence interactions between transcription factors and target genes in gene regulatory networks. A recent model called HoloNet innovatively applied the attention mechanism to decode the functional CCC events that have a critical effect on gene expression [18]. On this basis, we defined the concept of 'dominant cell communication assembly (DCA)' for the first time to denote cellular communication assemblies that dominate a specific downstream functional event. Then, we proposed scDCA, a multi-view graph learning method to decipher the DCA from scRNA-seq data. (i) Take advantage of four state-of-the-art CCC analysis methods, scDCA systematically and reliably infers L-R interactions from scRNA-seq data. (ii) According to prior knowledge of L-R interactions, gene expression profiles and cell type information, scDCA constructs the multi-view CCC network between different cell types at single-cell resolution by using an edge weighting strategy and filtering out edges with low specificity. (iii) scDCA develops a multi-view graph convolution network to reconstruct the expression pattern of target genes or the functional status of receiver cells and then deciphers the DCA by interpreting the trained model. In a scRNA-seq cohort of advanced renal cell carcinoma, scDCA was applied to decipher the DCA that affect the expression patterns of the critical marker genes of CD8+ T cell and tumor-associated macrophages, which accurately reflected the dominant cell regulation process on these cell types. scDCA also accurately deciphered the DCA that affect different functional states of malignant cells. Additionally, scDCA was applied to explore the alteration in cell communication under clinical intervention by comparing the DCA for certain cytotoxic factors between patients with and without immunotherapy. In summary, scDCA provided a valuable and practical tool for deciphering CCCs from scRNA-seq data, which helped to reveal biologically meaningful cell communication events.

## Material and methods

### Data preprocessing

The scRNA-seq data of patients with advanced renal cell carcinoma were derived from the recently published study [19] and accessed via the Single Cell Portal. The entire dataset contained 34 326 cells from tumor tissues of eight patients sequenced by 10X genomics' Visium platform. Cell samples from two patients (P76 and P915) were specifically selected. Compared to the untreated patient P76, patient P915 received immune checkpoint blockade (ICB, aPD-1 + aCTLA-4) and exhibited a clinical response. Both of them were not treated with tyrosine kinase inhibitor (TKI) and had the same biopsy Site, histology, and tumor stage for the subsequent comparison analysis (Table 1). The data preprocessing steps and cell type annotation analysis were referred to the original study. Specifically, a standard normalized unique molecular identifier count analysis workflow, including ambient RNA-decontamination with SoupX, library-size normalization, and log-transformation with Seurat, was applied. Uniform manifold approximation and projection (UMAP) of malignant and non-malignant cells captured across all lesions, colored by broad cell type. Granular cell types and states were discerned through iterative reprojection and unsupervised clustering of lymphoid, myeloid, and tumor compartments, and merged into broader cell type categories for visualization.

### Inference of L-R interactions

Previously, systematic assessment and comparison analyses showed that there is a dynamics of results among different L-R interactions inference methods [20, 21]. In addition, four of the statistics-based methods showed relatively higher performance [21], they are CellPhoneDB [6], CellChat [7], NicheNet [8], and ICELLNET [22]. To obtain the more comprehensive and reliable L-R interactions between every two different cell types for the following analysis, we intend to integrate the inferred results of the above four tools. In specific, the L-R interactions supported by at least two of the four methods were retained. Take the scRNA-seq data from patient P76 as a case study, we found the difference in the number of inferred L-R interactions by four distinct tools (NicheNet does not consider the interactions with multi-subunits, Supplementary Fig. 1a). Also, the overlap among the four methods showed the variation of the four methods (Supplementary Fig. 1b).

### Construct multi-view CCC graphs at the single cell resolution

#### Step 1: calculation of the communication probability of a specific L-R interaction

To calculate intercellular communication probability, we modeled L-R mediated signaling interactions using the law of mass action [23]. Based on the projected scRNA-seq profiles of ligands and receptors, the communication probability  $P_{ij}$  from cell  $i$  to  $j$  for a particular L-R pair  $k$  was modeled by

$$P_{ij}^k = \frac{L_i R_j}{K_h + L_i R_j} \quad (1)$$

$$L_i = \sqrt[m_1]{L_{i,1} \cdots L_{i,m_1}}, R_j = \sqrt[m_2]{R_{j,1} \cdots R_{j,m_2}} \quad (2)$$

where  $L_i$  and  $R_j$ , respectively, represent the expression level of ligand  $L$  in cell  $i$  and receptor  $R$  in cell  $j$ . The expression level of ligand  $L$  with  $m_1$  subunits (i.e.  $L_{i,1}, \dots, L_{i,m_1}$ ) is approximated by their geometric mean, implying that the zero expression of any subunit leads to an inactive ligand. Similarly, we compute the expression

level of receptor  $R$  with  $m_2$  subunits. In addition, a Hill function is used to model the interactions between  $L$  and  $R$  with a parameter  $K_h$  whose default value is set to be 0.5 as the input data has a normalized range from 0 to 1.

### Step 2: filtering out L-R interactions with low specificities

Considering the characteristic of graph learning, edges of non-specifically widely expressed ligands and receptors might introduce inappropriate embeddings. To reserve the L-R interactions that actively communicating, we used a permutation test strategy to calculate the specificity of each L-R interaction. Specifically, we first randomly selected  $n$  background gene pairs ( $n = 200$  by default) for each L-R pair. Note that the two genes of each background gene pair, respectively, have the most close average expression levels to the ligand and receptor. Then we calculated the communication probability of the background gene pairs to generate the null distribution of each L-R interactions. Communication probability of each background gene pair  $t$  between cell  $i$  and  $j$  for a particular L-R pair is

$$P'_{ij} = \frac{L_i^{(t)} R_j^{(t)}}{K_h + L_i^{(t)} R_j^{(t)}}, t = 1, \dots, n \quad (3)$$

where  $L_i^{(t)}$  and  $R_j^{(t)}$ , respectively, represent the expression level of  $t$ th background ligand and receptor in cell  $i$  and  $j$ . The communication probability of  $n$  background gene pairs  $P'_{ij}^{(1)}, \dots, P'_{ij}^{(n)}$  form a null distribution for the particular L-R pair. Finally, the L-R interactions whose communication probability is not over 95 percent of the null distribution will be filtered out. Supplementary Fig. 2 showed the proportion of the altered edges in the CCC network after filtering out L-R pairs with low specificities for the P76\_scRNA dataset.

### Step 3: edge weight setting in each view of network

According to the prior knowledge, the communication probability (edge weight in the network)  $w_{ij}$  between sender cell  $i$  and receiver cell  $j$  is calculated by summing up the communication probability of the filtered  $m$  L-R pairs between them:

$$w_{ij} = \sum_{k=1}^m p_{ij}^k \quad (4)$$

### Step 4: construction of the entire multi-view graph

The multi-view CCC network ultimately consists of networks under different combinations of cell types, e.g. cells between cell type A and cell type B form  $G_{A,B}$ , cells between cell type B and cell type C form  $G_{B,C}$ , etc., where the edge weights in the network are  $w_{ij}$ .

### Selection of target genes to be predicted

The selection of specific target genes was based on the following criteria: (i) they should not be genes of the mitochondrion, ligand, or receptor; (ii) they should be expressed in over 50% cells; (iii) they should be highly variable genes (detected by the `highly_variable_genes` function in Scanpy [24] package with default parameters).

### Predicting the specific target gene expression

To distinguish the regulatory effects of CCC on gene expression and the baseline expression levels within the corresponding cell

types, we hypothesized that the target gene expression  $E$  in all single cells could be separated into two parts:

$$E = E_0 + \Delta E \quad (5)$$

where the  $E_0$  is the baseline expression levels determined by its cell type, and  $\Delta E$  represents the gene expression change caused by the regulatory effects of CCC. Based on this, scDCA constructed two appropriate models to reconstruct target gene expression profile.

First, a graph convolution network is constructed for predicting  $\Delta E$ . This model can be broadly divided into three parts: (i) multi-view graph convolution network was applied to generate the embeddings of nodes (ii) the embeddings of nodes from each view were fused based on an attention mechanism (iii) multi-layer perceptron (MLP) model was applied to the final embeddings for predicting  $\Delta E$ .

Specifically, the inputs to the model include the adjacency matrix  $A = \{A^1, A^2, \dots, A^k\}$  of the multi-view communication network and the initial feature matrix  $X$  of the nodes (cells). The adjacency matrix  $A^k \in R^{N \times N}$ , where  $k$  represents the CCC network for the  $k$ th combination of cell types,  $N$  represents the number of cells, and the elements in the matrix  $A^k_{ij}$  represent the edge weights (communication probability)  $w_{ij}$  between cell  $i$  and  $j$ . The initial feature matrix  $X \in R^{N \times M}$  is generated by the one-hot encoding method based on cell type, where  $M$  represents the number of cell types. To obtain the embedding of nodes in each view, the GCN model [25] used each matrix in  $A$  as the adjacency matrix and  $X$  as the initial feature matrix. Then, the final embeddings of nodes were obtained by applying the attention mechanism to integrate the embeddings from each view. Finally, the MLP was used to predict  $\Delta E \in R^N$ :

$$\Delta E = \text{Linear} \left( \text{ReLU} \left( \sum_k c_k (A^k X W^k + B^k) \right) \right) \quad (6)$$

where  $c_k$  represents the learnable attention score,  $W^k$  represents the weight matrix,  $B^k$  represents the bias matrix of  $k$ th view, and  $\text{ReLU}$  is the activation function:

$$\text{ReLU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (7)$$

Second, the MLP model is constructed for predicting  $E_0$  based on the cell type matrix  $X$ :

$$E_0 = \text{Linear}(X) \quad (8)$$

Then the above two separate models jointly predict the target gene expression  $\hat{E}$ :

$$\hat{E} = \text{Sigmoid}(\Delta E + E_0) \quad (9)$$

where  $\text{Sigmoid}$  is the activation function:

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (10)$$

The mean squared error (MSE) between  $\hat{E}$  and true expression profile  $E$  was used as the loss function in this work:

$$\text{Loss} = \text{MSE}(E, \hat{E}) = \frac{1}{n} \sum_i (e_i - \hat{e}_i)^2 \quad (11)$$

## Calculation of the functional states activity of malignant cells

To characterize the functional states of each malignant cell, we obtained the signature gene sets of 14 crucial functional states of cancer cells, including stemness, invasion, metastasis, proliferation, epithelial-mesenchymal transition (EMT), angiogenesis, apoptosis, cell cycle, differentiation, DNA damage, DNA repair, hypoxia, inflammation, and quiescence from the CancerSEA database (<http://biocc.hrbmu.edu.cn/CancerSEA/>) [26]. Based on these signatures, the activities of 14 functional states across malignant cells were calculated using Gene Set Variation Analysis (GSVA) with the GSVA package in R.

## Predicting the functional state of the receiver cell

With a similar hypothesis to predicting target gene expression, the functional state activity  $F$  of the receiver cell was also separated into two parts:

$$F = F_0 + \Delta F \quad (12)$$

where the  $F_0$  is the baseline functional state activity for each receiver cell, and  $\Delta F$  represents the change of functional state activity caused by the regulatory effects of CCC. Also, the scDCA integrated the graph convolution network and MLP models to predict the functional state activity of receiver cells, similar to the construction process in predicting target gene expression. The only difference is that we only consider the functional state alteration of receiver cells (malignant cells here), and then the input adjacency matrixes (views of CCC graph) were limited to the communication between every non-malignant cell type and malignant cells. Furthermore, the initial feature of GCN is randomly initialized instead of the one-hot coding.

## Training strategy

In the study, we divide the dataset into training (85% cells) and validation sets (15% cells) of cell samples. For the predicted gene expression values in the training cells, the trained model directly predicts these values based on the multi-view CCC graphs and the cell type information. The entire training process is performed for 500 epochs, and at the end of each epoch, the trained model predicts the validation set, recording the mean-square error (MSE) between the true gene expression and the predicted gene expression of the validation set. Finally, the trained model under the epoch with the smallest MSE in the validation set is saved for interpretation. The Adam optimizer (initial learning rate of 0.1 and weight decay of  $5 \times 10^{-4}$ ) and the StepLR learning rate decay method (step size of 10 and gamma of 0.9) was, respectively, used. We are trying to adequately fit the gene expression of the validated cells, so as to obtain the attention values of different cellular communication views. In addition, to ensure the robustness of the model and the reliability of the interpretation, we performed 50 random divisions for each data set, and a total of 50 iterations of the model were trained. The attention score used for prioritization was the average value of 50 repeatedly trained models.

## Model interpretation for identifying the DCA

As mentioned above, we prioritized the views based on the attention scores that represent the contribution of the corresponding cell communication assembly on downstream gene expression or cell functional state. The attention score used for prioritization was the average value of 50 repeatedly trained models. The top-ranked cell communication assembly was defined the DCA. The

ratios of  $\Delta E/(\Delta E + E_0)$  and  $\Delta F/(\Delta F + F_0)$  reflected the extent that gene expression or cell functional state is affected by cell communication. The ratio is closer to 1, the more regulatory effect from cell communication on gene expression and cell function process.

## Collection of expression and clinical information from other independent ccRCC cohorts

Bulk RNA-seq samples of the kidney renal clear cell carcinoma (KIRC) from The Cancer Genome Atlas (TCGA) ( $N = 607$ ) was obtained. The normalized FPKM data was collected through the UCSC Xena database of the GDC project (<https://xena.ucsc.edu/>). Another two independent ccRCC cohorts (here denoted as Motzer\_NatMed\_2020 and Braun\_NatMed\_2020\_Che-ckmate025) with the treatment condition information were obtained from previous analysis to compare the prognostic effects of L-R pairs between immunotherapy and targeted therapy subpopulations. Normalized RNA-seq, treatment condition, and progression-free survival (PFS) data were collected from cohort NCT02684006, which incorporated with first-line avelumab + axitinib (anti-PD-L1 with TKI) vs sunitinib (multitarget TKI) in advanced renal cell carcinoma [27]. Also, we obtained bulk RNA-seq and overall survival data of pre-treatment samples from advanced ccRCC in a randomized clinical trial (Checkmate 025) comparing the mTOR inhibitor everolimus with nivolumab (anti-PD-1) [28].

## Statistics analysis

For comparisons of the continuous value between different groups, a two-sided Wilcoxon rank-sum test was performed. Co-expression of genes in TCGA samples was assessed based on Pearson's correlation analysis. For the survival analysis, Kaplan-Meier survival curves were generated and the log-rank test was used to determine the significant differences among patient groups. The hazard ratio and the 95% confidence interval were calculated using univariate Cox regression analysis. All statistics analyses were performed in R version 4.2.2.

## Functional enrichment analysis

Collectively, the g:Profiler tool (<https://biit.cs.ut.ee/gprofiler/gost>) was used to perform the functional enrichment analysis of Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The results of the enrichment analysis were visualized with the dot plot. To compare the genes more affected or less by CCC, we selected 5867 highly variable genes as targets and ranked them based on the performance improvements after considering CCC, i.e. difference of the Pearson's correlation coefficient between the predicted expression considering CCC effects or not and the real expression level). The top-ranked and bottom-ranked 50 genes were, respectively, selected for GO-biological process (GO-BP) enrichment analysis.

## Results

### The overview of scDCA workflow

The goal of scDCA is to decipher the DCA, i.e. CCCs with a dominant influence in a specific downstream functional event in the tumor microenvironment (TME). scDCA requires two inputs: a gene-by-cell count matrix and a cell type label vector. The main application scenarios include the following: (1) to decipher the DCA that affect the expression pattern of target genes (2) to decipher the impact of CCC for different cell types (3) to decipher the DCA that affect the functional states of receiver cells (4) to explore the alteration of DCA in cell communication



under clinical intervention. The workflow of scDCA involve four successive steps (Fig. 1), as briefly described below. More technical details are given in Methods.

### Step 1: construction of multi-view CCC network

scDCA is based on a multi-view CCC network between different cell types at single-cell resolution (Fig. 1a). First, four excellent CCC analysis tools were integrated to infer the L-R pairs from the scRNA-seq expression profile (see section 3 for details). Second, L-R interactions with low specificities would be filtered out and then the communication probability between a pair of single cells was calculated based on these remaining L-R interactions (see section 3 for details). Finally, the multi-view CCC network between each pair of different cell types was constructed based on the CCC adjacency matrix, where the nodes in the graph represent individual cells and the edges represent the communication probability between them.

### Step 2: decomposition of gene expression and cell functional state

Briefly, cell signaling from the sender cell is transmitted to the receiver cell and elicits downstream responses, such as alteration of gene expression or even the cell states of the receiver cell. Therefore, the expression ( $E$ ) of the specific target gene in the receiver cell can be defined as the composite of altered gene expression regulated by CCC ( $\Delta E$ ) and the baseline expression levels inherent in receiver cell type ( $E_0$ ). Similarly, we defined the functional state of the receiver cell ( $F$ ) be the composite of the changed functional state caused by the regulatory effects of CCC ( $\Delta F$ ) and the baseline functional state activity for the receiver cell ( $F_0$ ).

### Step 3: reconstruction of gene expression and cell functional state

scDCA constructed two models to reconstruct the ( $\Delta E$  or  $\Delta F$ ) and ( $E_0$  or  $F_0$ ) (Fig. 1c). For predicting the  $\Delta E$  or  $\Delta F$ , each view of the CCC graph was trained by a graph convolution network, node embeddings of each view were integrated through the attention mechanism and then retrained using a multi-layer perceptron (MLP). For predicting the  $E_0$  or  $F_0$ , another MLP was directly applied based on the cell type information or individual cell information. Finally, two estimations were summed as the final prediction results ( $\hat{E}$  or  $\hat{F}$ ). The model was trained iteratively to minimize the MSE between  $\hat{E}$  or  $\hat{F}$  and true expression profile ( $E$  or  $F$ ) (see section 3 and 3 for details).

### Step 4: interpretation of trained model

After the model converges, these different views were prioritized based on the attention weights of scDCA which represented the contribution of the corresponding cell communication assembly on downstream functional events (Fig. 1c). To ensure the reliability of the results, we repeated the training procedure and used the average attention value as results. The communicating cell type pair corresponding to the higher ranked views were regarded as the DCAs (Fig. 1d).

## Resolving the CCC at single-cell resolution is necessary

To verify the necessity of analyzing CCC at single-cell resolution, we inferred the intercellular communication status of an advanced renal cell carcinoma microenvironment dataset from a specific patient P76 (see section 3 for details). We inherited the

cell type annotation from the data resource analysis [19], and performed the UMAP analysis to verify the different cell populations (Supplementary Fig. 3a). We found that malignant, TAM, NK, and CD8+ T cells had greater proportion (Supplementary Fig. 3b).

Then, we demonstrated the heterogeneity of CCC between different cells at the single-cell resolution from different perspectives. We randomly selected 100 immune cells (50 TAM cells and 50 CD8+ T cells) and 100 malignant cells, and then calculated and compared the communication strength between the two types of immune cells and malignant cells at the single-cell resolution (see section 3 for details). It is obvious from the heatmap in Supplementary Fig. 3c that for the same malignant cell, different types of individual immune cells send out communication signals with different strengths (the color shades in the heatmap represents the strength of CCC between two individual cells). In addition, even for individuals of the same immune cell type, their communication strength with malignant cells varies dramatically (The histogram in Supplementary Fig. 3c represents the total communication strength of individual immune cells with malignant cells). Supplementary Fig. 3d shows the distribution of the number of communicational L-R pairs between immune and malignant cells in Supplementary Fig. 3(c). In particular, most of the cell pairs communicated through approximately one L-R pair, very few cell pairs produce more than four L-R pairs-mediated CCC. Besides, we further compared the top L-R pair with the highest communication strength between immune and malignant cells in Supplementary Fig. 3c. The color shades in Supplementary Fig. 3e represent the type of the top L-R pair, again showing a high degree of heterogeneity. Finally, we counted the proportions of different L-R pairs expressed in all cell pairs between TAMs or CD8+ T cells and malignant cells. Unsurprisingly, either TAMs or CD8+ T cells exhibited diverse landscapes of L-R pairs mediating CCC with malignant cells (Supplementary Fig. 3f).

In summary, CCCs are highly heterogeneous at the single-cell resolution. By considering CCCs at the single-cell resolution rather than at the cell type level, signals from different cellular sources can be integrated to provide a more comprehensive picture of the real CCCs and downstream functional events in the tumour microenvironment.

## scDCA reveals the DCA affects the marker genes of CD8+ T cells

We applied scDCA to scRNA-seq data from a specific patient with advanced renal cell carcinoma as a case study (P76\_scRNA, see section 3 for details). Specifically, we constructed a multi-view CCC network, which connects 7393 single cells of 13 cell types, resulting in 169 views that systematically depict the CCC in the TME. The cell types include CD8+ T cells, T helper cells (T-Helper), regulatory T cells (T-Reg), NK cells, NKT cells, B cells, plasma cells, dendritic cells (DC), TAM, monocyte, cycling cells, malignant cells, and endothelial cells (Supplementary Fig. 3a and b) [19].

To demonstrate the validity of scDCA, we first assessed its performance to reconstruct the target gene expression. We selected CD8A (CD8 Subunit Alpha) as an example (Fig. 2a). It codes for a subunit of cell surface glycoprotein CD8 commonly found on most cytotoxic T lymphocytes (CD8+ T cells), and mediates efficient CCCs involving anti-tumor immune response. As shown in Fig. 2b, the reconstructed CD8A expression pattern by scDCA was highly consistent with the actual expression profile and showed a significant correlation (Pearson correlation  $r = 0.702$ ,  $p$ -value  $< 0.05$ ). The ratio of  $\Delta E$  to total expression levels ( $\Delta E + E_0$ ) reflected the degree to which the target gene is affected by CCC (see section 3 for details). That is, the higher ratio indicates the gene

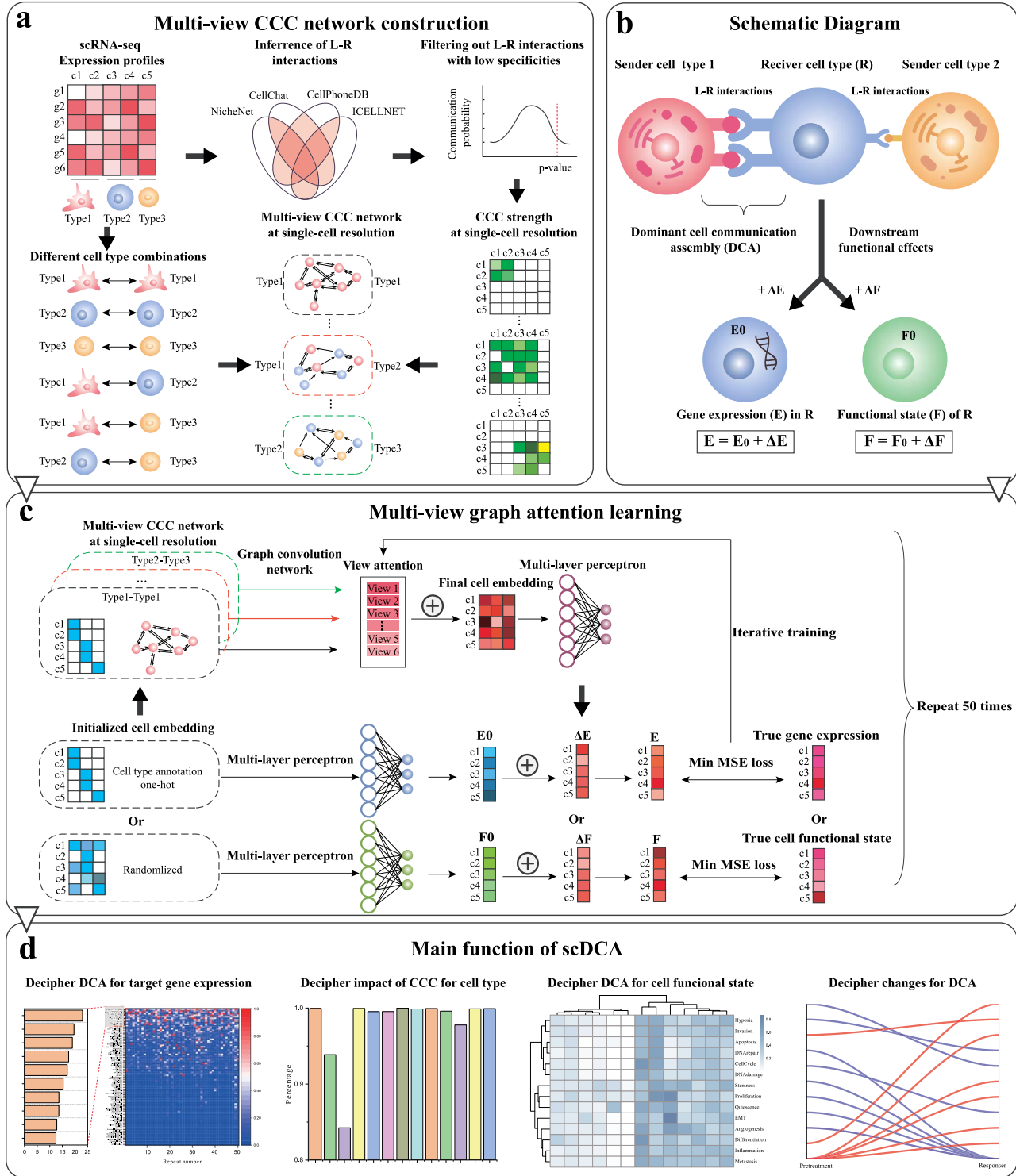


Figure 1. The overview of scDCA workflow. (a) The construction of multi-view CCC network at single-cell resolution. See section 3 for details. (b) The schematic diagram of a full picture of CCC and its downstream functional events. LR: ligand-receptor; E: gene expression profile;  $E_0$ : baseline expression determined by cell type;  $\Delta E$ : expression change caused by CCC. F: functional states activity of receiver cells;  $F_0$ : baseline state activity determined by the individual receiver cell;  $\Delta F$ : change of state activity caused by CCC. (c) The multi-view graph convolution network for reconstructing target gene expression patterns or functional states of receiver cells. MSE: mean squared error. See section 3 for details. (d) The main function of scDCA.

expression is more affected by CCC, and the lower ratio indicates the gene expression is more affected by intracellular regulation. In patient P76, the ratio of CD8A was calculated in each cell type. As shown in Fig. 2c, the expression of CD8A in CD8+ T cells and Cycling cells was less affected by CCC, which corresponds to the

specific high expression of CD8A in these cell types (Fig. 2a). In addition, we extended the analysis to other genes. The ratio was calculated for all of the 5867 highly variable genes (HVGs) in the dataset. In particular, scDCA reconstructed gene expression profiles with and without the addition of the CCC network

respectively. The degree of improvement in the reconstruction accuracy of the gene expression profile (as indicated by Pearson correlation

coefficients) after the addition of the CCC network as the degree to which the gene is affected by CCCs. Through GO enrichment analysis, the top 50 affected HVGs by CCC showed functional roles involving cell communication, cellular response to stimulus, chemical, and stress (Supplementary Fig. 4). On the contrary, the bottom 50 affected HVGs by CCC were enriched in immune cell-specific biological processes, such as T cell-mediated cytotoxicity, natural killer cell-mediated cytotoxicity, lymphocyte differentiation, mononuclear cell differentiation, MHC protein complex assembly, and cell development (Supplementary Fig. 4). The enrichment results showed that scDCA accurately deciphered the expression profiles of target genes.

Then we assessed the scDCA in prioritizing the DCA of target genes in receiver cells. Based on the attention value of the graph convolution network (see section 3 for details), scDCA ranked the contribution for each cell type pair (views). Also, take the CD8A as an example, results showed that its expression might mainly influenced by CD8+ T cell-Malignant interactions and Malignant-TAM interactions (Fig. 2d). As a marker gene of CD8+ T cells, DCA affecting CD8A expression might also reflect the primary communicating cell types of CD8+ T cells in the TME. Based on the previous analysis, we calculated a signature score based on the average expression of several regulatory genes of T cells, including CD70, ICOSLG, CD155 (PVR), CD112 (NECTIN2), and PD-L1 [29]. It is observed this regulatory score is higher in malignant cells, which was consistent with the results of scDCA (Fig. 2e). In addition, previous studies validated that malignant cells induced macrophages to express high levels of IL-15R $\alpha^+$ , which reduced the protein levels of chemokine CX3CL1 in malignant cells to inhibit the recruitment of CD8+ T cells [30]. Also, malignant cells with stem-like states were revealed to drive T cell dysfunction via immunosuppressive macrophages [31]. These findings all supported the influence of malignant and its interactions with TAM on CD8+ T cells, which is consistent with the results of scDCA.

To further compare the communication strength of CD8+ T cells and malignant/TAM cells with that of CD8+ T cells and other cell types in the TME, we calculated the average communication probability with all CD8+ T cells for each single cell. The comparison supported that malignant and TAM cells had significantly higher communication strength with CD8+ T cells than the rest cell types (Fig. 2f). In terms of each cell type, TAM still had the highest communication strength, while malignant cells had a lower median value but a wider range (Fig. 2g). For malignant cells, we further divided them into two groups based on their average communication strength with CD8+ T cells. The group with higher communication strength had accordingly higher activity of antigen presentation via MHC I and IFN- $\gamma$  signaling (Supplementary Fig. 5). These observations suggested strong cellular communication between a subset of malignant cells and CD8+ T cells. All of the above investigations revealed the dominant CCC between CD8+ T and malignant/TAM cells which again supported the results of scDCA.

### scDCA reveals the DCA affects the key regulatory genes in tumor-associated macrophages

In section 3, we found that TAMs have strong intercellular communications with CD8+ T cells, which made us interested in the DCA associated with TAMs. We further deciphered several key regulatory genes in TAMs, including FOLR2 (Folate receptor  $\beta$ ), APOE (Apolipoprotein E), and FABP5 (Fatty Acid Binding Protein

5) [32–34]. A multi-view CCC network centered on the TAMs was constructed, resulting in 13 views that involved the communications between other 12 cell types and TAMs as well as TAM themselves. Then, scDCA was retrained and priority sorted on dominant cellular communication based on attention weight in anticipation of revealing holographic networks for specific TAM markers. After training, scDCA similarly accurately reconstructed the gene expression profiles, which were highly consistent with the real gene expression profiles (Supplementary Fig. 6).

According to the analysed results, the expression of FOLR2 in TAMs was mainly affected by the communication between malignant cells and TAMs, while APOE and FABP5 were greatly affected by the communication within TAM cells (Fig. 3a). FOLR2 was canonically described as a marker of macrophages M2 polarization [32], and indeed specifically over-expressed in part of the TAMs (Supplementary Fig. 6a). In the TME, M2 polarization is usually associated with the tumor-promoting phenotype of TAMs [35]. APOE was found to be specifically over-expressed in TAMs in the previous analysis about gastric and breast cancers [33], consistently, we also observed its high expression in TAMs of our ccRCC dataset (Supplementary Fig. 6b). It also verified as a potential prognostic biomarker for ccRCC recurrence by [36]. Besides, previous analysis supported that FABP5 controlled macrophage alternative activation, also, FABP5-expressing TAMs were present at invasive cancer regions, produced various immunoregulatory molecules (including PD-L1 and PD-L2) [35, 37].

Further, taking FOLR2 as an example, we further analyzed it from the ligand-receptor-target (L-R-T) perspective. We integrated L-R-T regulatory data from CellPhoneDB and NicheNet to explore important L-R pairs that regulated FOLR2 expression in TAMs and the relationship between these L-R pairs and malignant cells. In the results, CellPhoneDB revealed L-R pairs of APP:TREM2, APP:SORL1, MDK:SORL1 and MDK:LRP1 significantly mediated the communication between malignant cells and TAMs as well as TAMs themselves (Fig. 3b). NicheNet evaluated the potential of different ligands to regulate FOLR2. The ligands APP and MDK were at the top of the list, especially APP was the most effective ligand in regulating FOLR2 (Fig. 3c). Specifically, ligands APP and MDK were highly expressed in the malignant cells while the corresponding receptors TREM2, SORL1, and LRP1 were highly expressed in the TAMs (Supplementary Fig. 7), therefore, they might jointly mediate the regulation of FOLR2 expression in TAMs (Fig. 3d). In addition, co-expression analysis in the bulk RNA-seq samples from TCGA also supported the regulation of these L-R pairs on FOLR2 (Supplementary Fig. 8). It is worth noting that APP and MDK regulate different receptors in different directions, which in turn results in a superposition of multiple effects on FOLR2. All the results demonstrate the accuracy and reliability of scDCA for deciphering the DCA that regulates specific target genes and facilitates further effective excavation of mediating L-R pairs.

### scDCA reveals the DCA affects the functional states of malignant cells

Compelling evidence highlighted that multiple CCCs between different cell types were involved in affecting functional states of cells, such as angiogenesis [38], EMT [39], and invasion [40] of malignant cells. scDCA could be also used to decipher the impact of each communicating cell type pairs on the cell states. To do this, we first calculated the activity scores for 14 functional states of 7393 malignant cells in our P76\_scRNA data. There was a high degree of functional heterogeneity among these malignant cells, with a small fraction of them exhibiting elevated levels

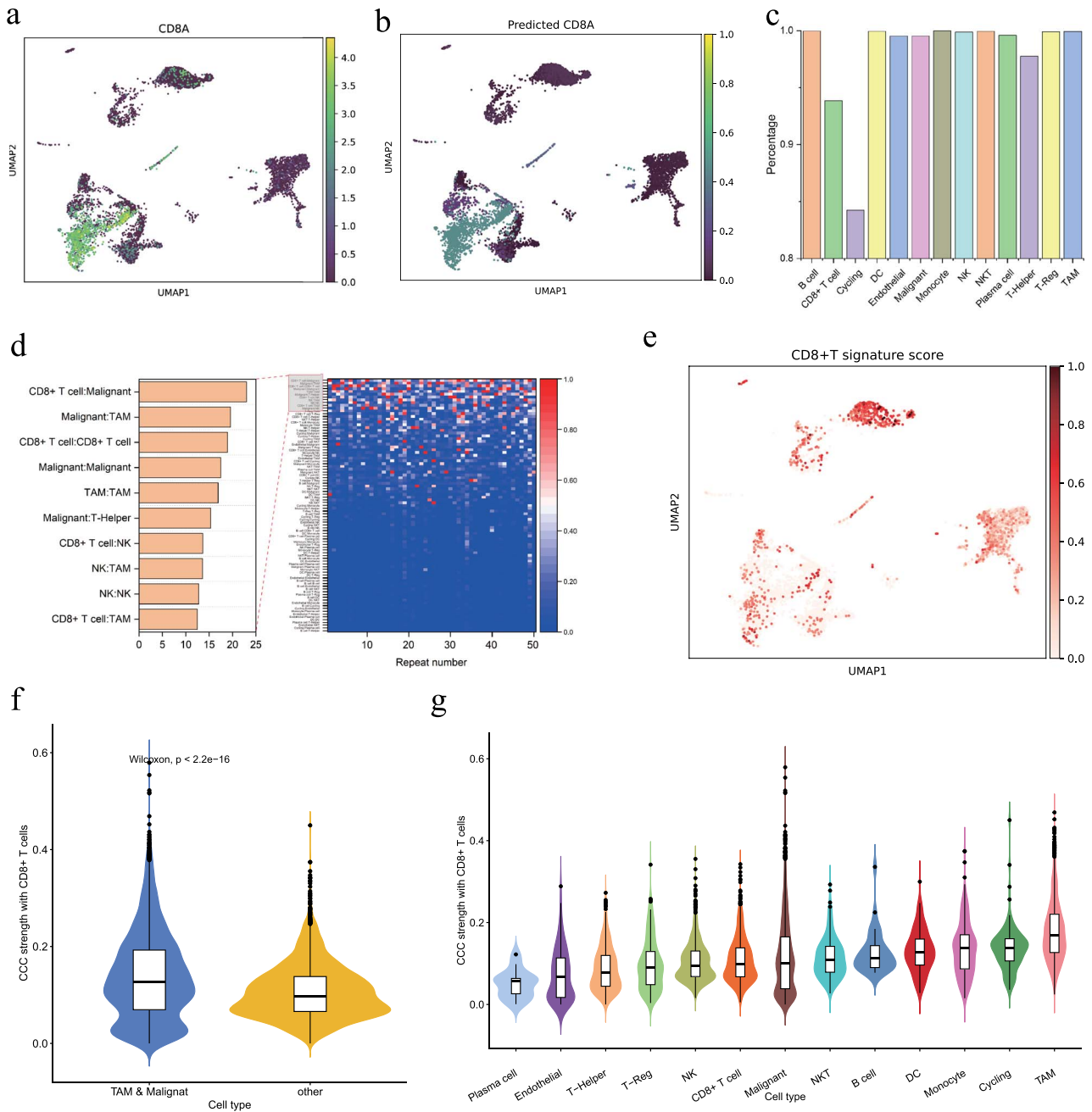


Figure 2. The DCA that regulates the CD8A expression pattern is accurately deciphered. (a) UMAP of cell samples, followed by the real expression pattern of the CD8A gene in UMAP space. (b) UMAP of cell samples, followed by the predicted expression of CD8A gene by scDCA. (c) The ratio of  $\Delta E$  to the total expression levels ( $\Delta E$  and  $E_0$ ) in each cell type. (d) The corresponding cell type pairs of the top 10 view (169 views in total) with the highest attention weights in the model predicting CD8A expression profile. The heatmap displays the attention weights of each view obtained from repeated the training procedure 50 times. The bar plot represents the mean values of the attention weights of each view. (e) UMAP of cell samples, followed by the CD8+ T signature score in UMAP space. (f) Comparison of communication strength of CD8+ T cells and malignant/TAM cells with that of CD8+ T cells and other cell types. (g) Comparison of communication strength with CD8+ T cells among each cell types.

of EMT, invasion, metastasis, and hypoxia, suggesting a higher malignancy potential in this subpopulation (Supplementary Fig. 9). Then, scDCA was retrained based on the reconstructed multi-view CCC network centered on the malignant cells, resulting in 13 views that involved the communications between 12 other cell types and malignant cells as well as malignant cells themselves. Different from the gene expression analysis in the previous section, our goal is to predict the functional state scores of malignant cells and assign initial random feature vectors to each individual malignant cell due to the fact that the functional state is

associated with the individual cell (see section 3 for details). In the results, scDCA had significant accuracy in predicting various cell functional states, such as the DNA repair state with a Pearson correlation of 0.875 and p-value  $< 0.05$ , and the apoptosis state with a Pearson correlation of 0.879 and p-value  $< 0.05$  (see Supplementary Table.2 for more details). Similarly based on the average attention weight of 50 repetitions of training, scDCA deciphered the DCA that affected the functional state of malignant cells as shown in Fig. 4a. Overall, natural killer (NK), T-helper, malignant itself, CD8+ T, TAMs, endothelial and monocyte cells



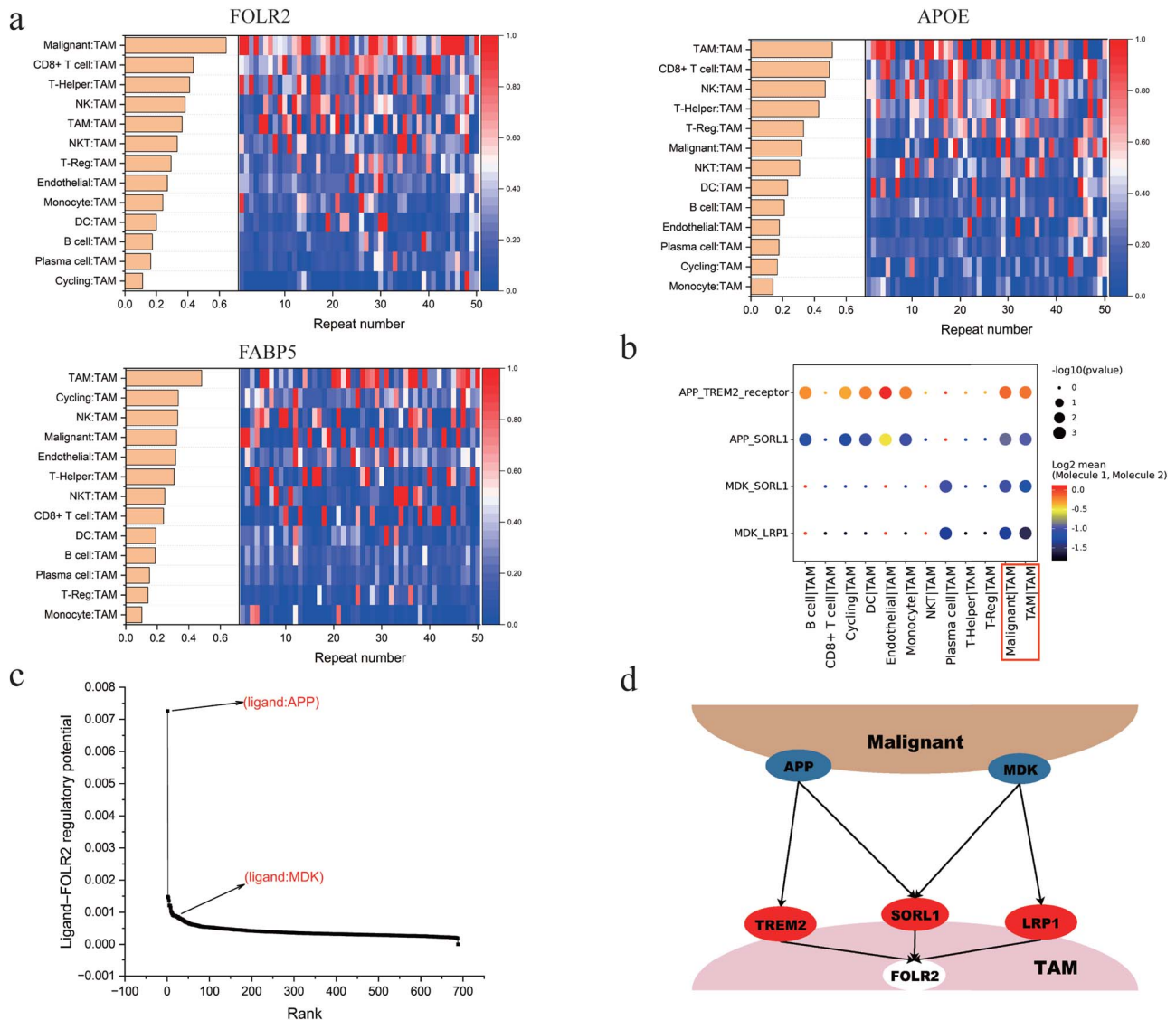


Figure 3. The DCA that affects the key regulatory in tumor-associated macrophages is accurately deciphered. (a) Rank of corresponding cell type pairs (views, 13 in total) based on the attention weights in the model predicting expression profile of the key factors FOLR2, APOE, FABP5 in TAMs, respectively. The heatmap displayed the attention weights of each view obtained from repeating the training procedure 50 times. The bar plot represents the average attention weights of each view. (b) The significant L-R interactions between malignant cells and TAMs which assessed by the CellPhoneDB analysis. (c) Rank of ligands based on the regulatory potential on target gene FOLR2 calculated by the priori database NicheNet. (d) The schematic diagram of ligands APP, and MDK from malignant cells bind to the receptors TREM2, SORL1, and LRP1, respectively, to coordinately regulate the expression of FOLR2 in TAMs.

had a greater impact on the functional states of malignant cells. In contrast, B and plasma cells had less overall effect. From the results, the cell states that were significantly influenced by the malignant cells themselves were angiogenesis and differentiation. Indeed, we observed VEGFA (vascular endothelial growth factor), a major signal that influences angiogenesis and promotes differentiation of malignant cells [41], is highly specific expression in malignant cells (Fig. 4b). In addition, NK cells affected various functional states of malignant cells, especially their cell cycle and differentiation. This phenomenon has been shown by recent studies in which NK cells upregulated the expression of specific genes to induce cell cycle or differentiation of malignant cells [42–44], which further corroborates the results of scDCA.

Furthermore, tumor invasion and metastasis involve a sophisticated cascade process, often referred to as the 'invasion-metastasis cascade', and EMT has been shown to be a critical

first step for the cascade process [45–47]. In our results, the CCC between malignant and endothelial cells was the most influential factor for the EMT state of malignant cells, this finding was strongly supported by the prior knowledge about the EMT process [48]. Considering the three cascade states comprehensively, we observed that endothelial cells, NK, T-helper, and TAM cells were the dominant cell types that were involved in regulating them (Supplementary Fig. 10). Intersection of the dominant cell types showed NK cells and TAMs might exert common effects for the 'invasion-metastasis cascade' of malignant cells, which had been validated by substantial studies. NK cells play a critical role in the control of metastasis of malignant cells as they mediate metastatic cell immunoediting, [49, 50], and IFN- $\gamma$ -dependent NK cell activation is vital to suppress malignant metastasis [51]. Also, perturbation of the EMT state of malignant cell lines altered several representative

NK cell-regulatory ligands [52]. Similarly, TAM cells was found to promote malignant progression by enhancing invasion and metastasis [53, 54]. Specifically, M2-polarized TAM was found to promote EMT and activate invasion programs in malignant cells [55]. The expression pattern of some specific metastatic regulators also supported the critical influence of NK and TAMs cells on the 'invasion-metastasis cascade' states of malignant cells (Fig. 4c). In particular, Apolipoprotein C-I (APOC1) has been identified as a novel pro-metastatic factor [56], which enhances the metastasis of malignant cells via EMT pathway in ccRCC and gastric cancer [57, 58]. Indeed, we observed that APOC1 was specifically highly expressed on TAMs in our data, which supports the potential of TAMs in promoting malignant cell metastasis as revealed by scDCA (Fig. 4c). Another example is the NK-derived chemokine CCL5, which promotes malignant metastasis and has been verified as an important mediator for immune evasion in circulating malignant cells [59]. Finally, we analyzed the correlation between 'invasion-metastasis cascade' signature genes and TAM/NK cell marker genes at the bulk level. We first filtered the three sets of 'invasion-metastasis cascade' associated genes based on the malignant cell-specific expressed genes. Their average expression was found significantly correlated in TCGA-KIRC samples ( $N = 607$ ) (Fig. 4d). These collective findings provide compelling evidence that scDCA possesses the capability to decipher intercellular communication that affects the functional status of malignant cells in the TME.

### scDCA reveals the altered DCA under the clinical intervention of immune checkpoint blockade

Understanding the alteration of DCA under therapeutic interventions is crucial for dissecting the specific molecular mechanism of disease progression or therapeutic response. To compare with the previously analyzed untreated patient P76, we obtained scRNA-seq data from another patient with advanced ccRCC, P915, who received ICB therapy and exhibited a clinical response [19] (Supplementary Table 1). Totally 6304 cells were obtained and 14 cell types were annotated according to the original publication (Supplementary Fig. 11a). We can see that there was a significant increase of CD8+ T cells after ICB treatment (Supplementary Fig. 11b).

The scDCA was then applied to decipher the DCA that affects CD8A expression pattern in ICB-responder P915 scRNA-seq data. Compared to the untreated patient (Fig. 2d), the top DCA changed from CD8+ T cell-malignant to CD8+ T cell-TAM (Fig. 5c). To further investigate the DCA alteration in the ICI-responders, we analyzed two additional genes, GZMA and PRF1, which are signature genes of cytolytic activity, indicating the ability of CD8+ T cells to clear malignant cells [60, 61]. Notably, similar to CD8A, DCA that affects these two cytolytic indicators also turned into CD8+ T cell-TAM (Fig. 5c). These findings were supported by the previous observation [19] that normalized effect or molecule (e.g. IFNG) levels were robustly correlated with TAM fractions and expression signatures derived from comparing ICB (immune checkpoint blockade)-exposed to ICB-naive TAMs. All these results suggested the increased potential of TAM in regulating CD8+ T cells in ICI-responders compared to untreated patient.

Additionally, the scDCA was also applied to decipher the DCA that affected three cellular signatures associated with the clinical response of ICB [19], including the tumor program 1 (TP1) and tumor program 2 (TP2) active in malignant cells that may drive interactions with the immune system, as well as the signature of immune checkpoint and evasion

(Supplementary Fig. 12). We found that the TAM-malignant cell pair significantly rose to the top 3 ranks related to treatment response in ICB-responders compared to untreated patient, suggesting that TAMs affected the therapeutic responsiveness of malignant cells.

Therefore, we speculated that the CCC changes generated by TAM after ICB therapeutic intervention are crucial in mediating immunotherapy response. On the basis, we then obtained the differential expressed genes of TAMs derived from comparing ICB-responder (P915) to the untreated patient (P76). Functional enrichment analysis of GO-biological process revealed the up-regulated genes after ICB treatment were related to several processes involving immune response and cell death, such as 'programmed cell death', 'apoptotic process', 'autophagy', 'inflammatory response', 'proliferation, differentiation and activation of T cell', and 'cytokine production' (Fig. 5a). In terms of pathways analysis, similar to GO-BP results, the immune system, antigen processing and presentation, cytokine signaling, and programmed cell death were significantly enriched. Besides, metabolic pathways and VEGFA:VEGFR2 signaling were also enriched (Fig. 5b). These results indicated the reinforced immune regulating potential of TAM cells in ICB-responder dataset.

Furthermore, the differences in L-R interactions between TAMs and CD8+ T cells were further compared. To our surprise, there was a significant increase in L-R pairs in the ICB-responder dataset compared to the untreated one. In specific, according to cellPhoneDB analysis, there were 62 L-R pairs between TAMs and CD8+ T cells in the untreated patient while 1236 L-R pairs in the ICB-responder patient, respectively. For the 1028 different LR pairs, we further investigated their correlation with the clinical benefit of the ICI treatment. Using the average expression of each LR pair, we found that 11 LR pairs exhibited significant prognostic significance in two independent advanced ccRCC cohorts following ICB immunotherapy (Supplementary Fig. 11c). Importantly, their prognostic association was significant only within subpopulations receiving ICB immunotherapy, compared to Sunitinib [27] and mTOR inhibitor everolimus [28] (Supplementary Fig. 13a and b). Furthermore, when combining the average expression of these 11 L-R pairs, a similar prognostic association was observed in ICB-treatment-only subpopulations but not the targeted therapy subpopulations (Fig. 5d, Supplementary Fig. 13c and d).

In summary, we attempted to apply scDCA in patients with different clinical phenotypes and explored the alteration in CCC that have a unique contribution to clinical intervention and patient prognosis. In the comparison of untreated and ICI-responder patients of advanced RCC, scDCA helped discern the prominent role of TAM in regulating the response of CD8+ T cells after immunotherapy and the emerged L-R pairs showed potential that mediating clinical benefit of immunotherapy.

## Conclusions

In response to the computational challenge of deciphering the downstream functional impact of CCC, we proposed a computational method, scDCA. It can innovatively prioritize cell type pairs that play a dominant role in downstream functional impacts (called DCA). In particular, scDCA builds the comprehensive structure of CCC at the single cell resolution to maximally leverage the information from communicative heterogeneity and specificity. We construct the multi-view graph convolution network based on the scRNA-seq data to model the impact of different cell

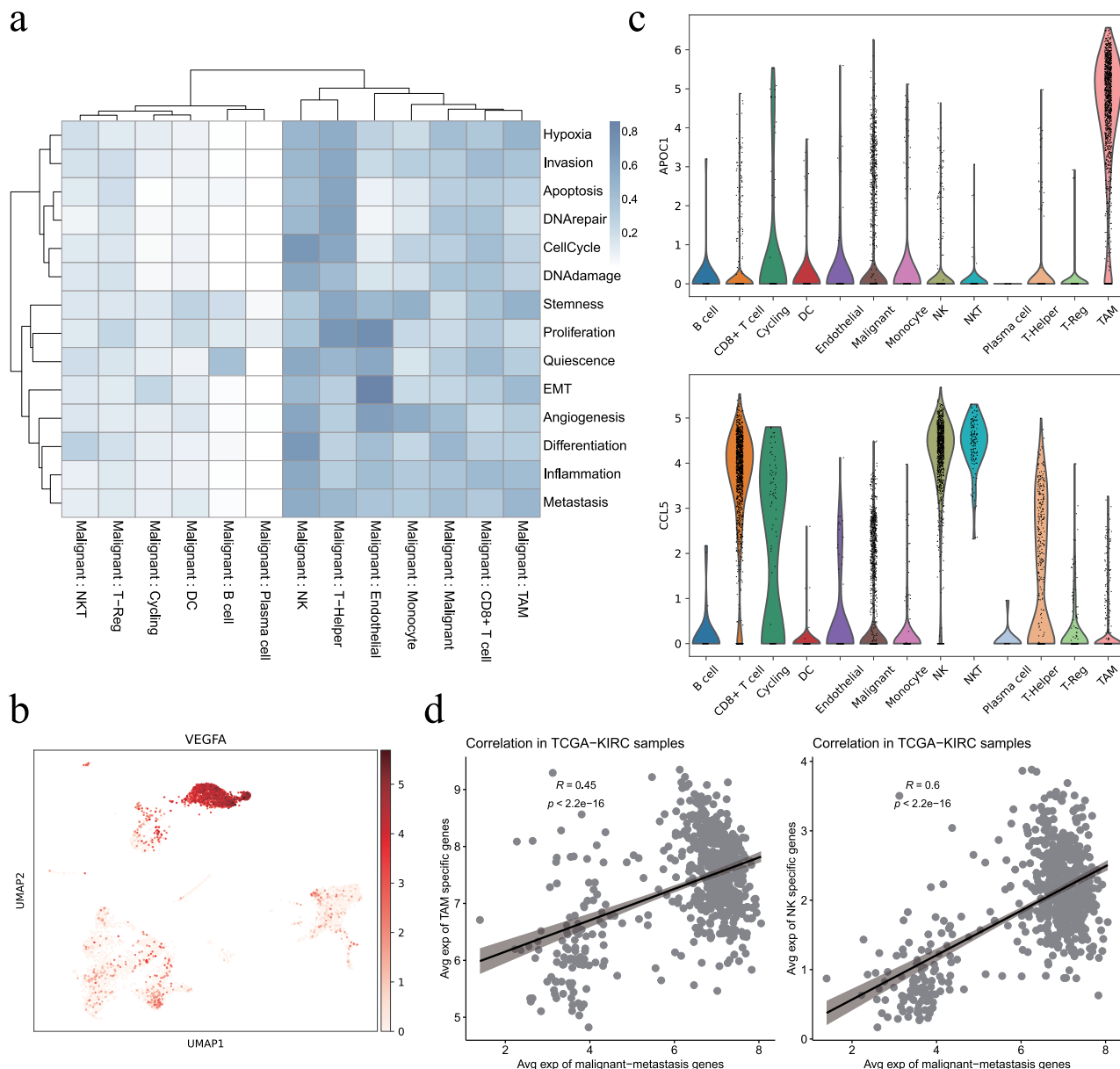


Figure 4. The DCA that affects functional states of malignant cells is accurately deciphered. (a) Rank of attention weights of corresponding cell type pairs (views) between malignant cells and other cell types in the model predicting the functional states of malignant cells. The heatmap displayed the attention weights of each view obtained from repeated the training procedure 50 times. The bar plot represented the average attention weights of each view. (b) UMAP of cell samples, followed by the expression pattern of VEGFA. (c) Expression distribution of the two metastatic regulators (TAM-specific gene APOC1 and NK-specific gene CCL5) in different cell types. (d) Correlation analysis between signature of 'invasion-metastasis cascade' states of malignant cells and marker genes of TAMs and NK cells in the TCGA-KIRC samples ( $N = 607$ ).

type pairs on the target gene expression or functional states of cells. In advanced renal cell carcinoma dataset, scDCA reveals the DCA that affects the key factors in CD8+ T cells and TAMs as well as the multiple functional states of malignant cells. In addition, it also provides insights into changes in DCA before and after the immunotherapy so as to facilitate understanding of the intercellular communication basis of the clinical response.

In the future, scDCA can be extended to incorporate single-cell multiomics and spatial RNA-seq data to enhance its predictive power for deciphering DCAs. By integrating molecular layers such as proteomics and epigenomics, scDCA could refine ligand-receptor interaction modeling and account for regulatory mechanisms underlying gene expression. Spatial RNA-seq data,

with their spatial coordinates, would allow scDCA to identify biologically plausible CCCs and construct spatially informed CCC graphs. These advancements would improve graph-learning approaches, enabling more precise modeling of CCC effects on target genes and functional states.

While scDCA offers robust insights into CCCs, it has limitations. First, scDCA is based on the construction of multi-view CCC graph, so its accuracy is inevitably constrained by the current completeness of the current ligand-receptor databases. Also, the pre-constructed CCC graph might limited this method in generalization for novel cell types or conditions. Then, considering the input data, single-cell RNA-seq data are usually sparse, so the effect of this sparsity should be further considered in the

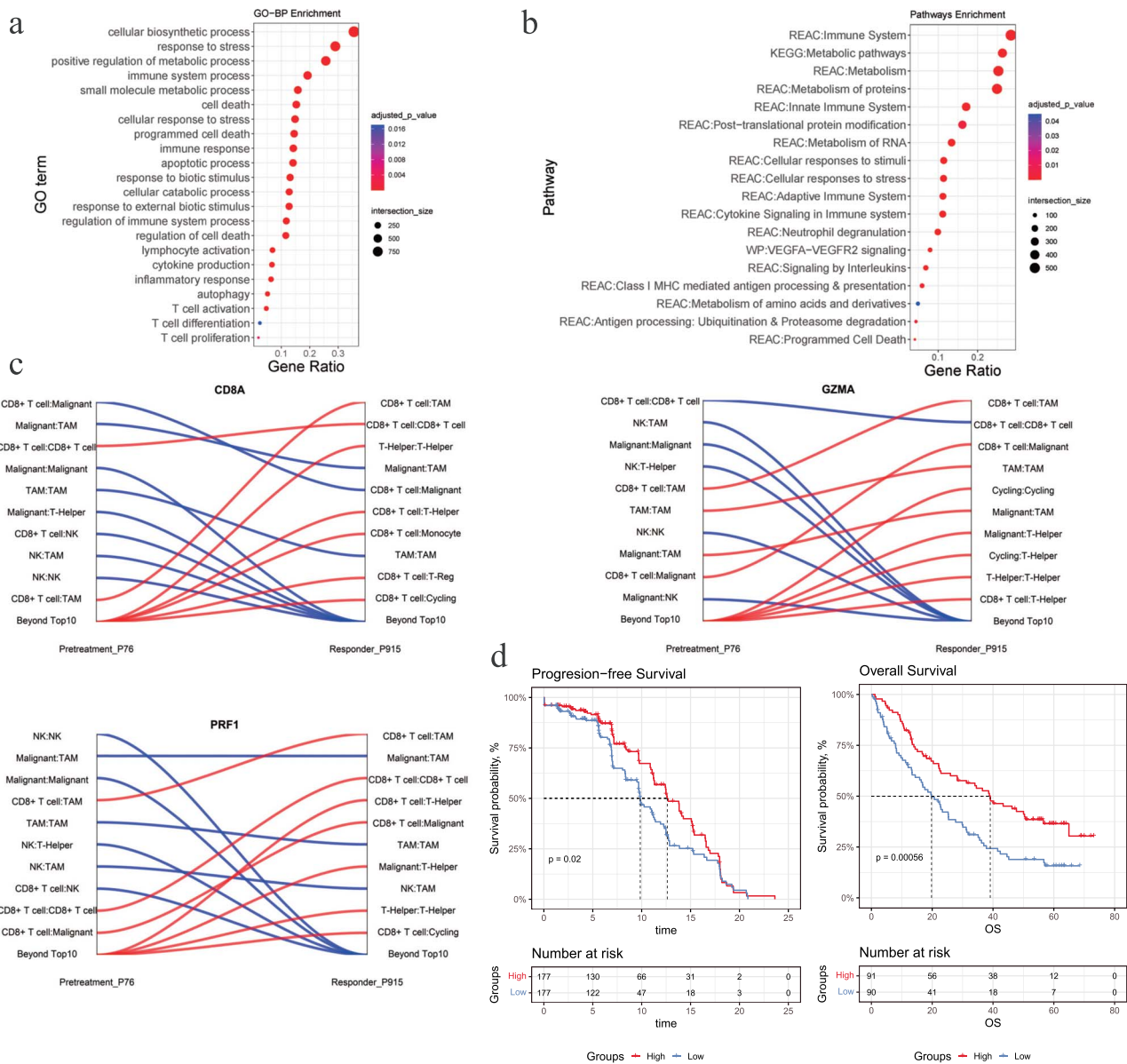


Figure 5. The DCA associated with clinical response of immune checkpoint blockade is accurately deciphered. (a) Functional enrichment analysis of GO-BP and KEGG pathways (b) for the differential expressed genes of TAMs derived from comparing the ICB-responder (P915) to the untreated patient (P76). (c) Mapping of the rank changes of DCA for three signature genes from comparing ICB-responder (P915) to untreated patients (P76). (d) Kaplan-Meier curves and log-rank test for the integration of the 11 L-R pairs in the two ICB-treatment-only subpopulations. The average expression of these LR pairs was summed up for the prognostic association analysis. Left: data from ICB immunotherapy-only subpopulations of Motzer\_NatMed\_2020 cohort; Right: data from ICB immunotherapy-only subpopulations of Braun\_NatMed\_2020\_Checkmate025 cohort.

reconstruction model of gene expression. In addition, multi-view graph learning of CCC network at single-cell resolution requires sufficient computational resources, which may limit its the scalability of scDCA. Addressing these limitations in future iterations could further enhance our method's applicability and accuracy across diverse biological contexts.

### Key Points

- As a new computational method, scDCA can be used to decipher the dominant cell communication assembly (DCA) of a particular functional event in receiver cells from single-cell RNA-seq data.

- scDCA efficiently utilizes multi-view graph convolution network and attention mechanism to reconstruct the cell-cell communication landscape and identifies DCA.
- scDCA is successfully applied and validated in revealing the DCA affecting the crucial gene expression in immune cells and the functional states of malignant cells in advanced renal cell carcinoma samples.
- scDCA is successfully applied and validated in exploring the alteration of cell-cell communications under clinical intervention.
- Our data, code, and a step-by-step tutorial are free available at a public GitHub repository: <https://github.com/pengsl-lab/scDCA.git>.



## Supplementary data

Supplementary data is available at Briefings in Bioinformatics online.

Conflict of interests: None declared.

## Funding

This work was supported by National Key R&D Program of China 2022YFC3400400, 2023YFC3503400; NSFC-FDCT Grants 62361166662; The Innovative Research Group Project of Hunan Province 2024JJ1002; Key R&D Program of Hunan Province 2023GK2004, 2023SK2059, 2023SK2060; Top 10 Technical Key Project in Hunan Province 2023GK1010; Key Technologies R&D Program of Guangdong Province (2023B1111030004 to F.F.H.). The Funds of State Key Laboratory of Chemo/Biosensing and Chemo-metrics, the National Supercomputing Center in Changsha, and Peng Cheng Lab; Graduate Research Innovation Project of Hunan Province (QL20230101).

## References

1. Armingol E, Officer A, Harismendy O. et al. Deciphering cell-cell interactions and communication from gene expression. *Nat Rev Genet* 2021;**22**:71–88. <https://doi.org/10.1038/s41576-020-00292-x>.
2. Zhang H, Lin X, Han Z. et al. Crystal structure of PXY-TDIF complex reveals a conserved recognition mechanism among CLE peptide-receptor pairs. *Cell Res* 2016;**26**:543–55. <https://doi.org/10.1038/cr.2016.45>.
3. He Y, Rodrigues RM, Wang X. et al. Neutrophil-to-hepatocyte communication via LDLR-dependent miR-223-enriched extracellular vesicle transfer ameliorates nonalcoholic steatohepatitis. *J Clin Invest* 2021;**131**:300–310. <https://doi.org/10.1172/JCI141513>.
4. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature* 2017;**541**:321–30. <https://doi.org/10.1038/nature21349>.
5. Kalluri R. The biology and function of fibroblasts in cancer. *Nat Rev Cancer* 2016;**16**:582–98. <https://doi.org/10.1038/nrc.2016.73>.
6. Efremova M, Vento-Tormo M, Teichmann SA. et al. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* 2020;**15**:1484–506. <https://doi.org/10.1038/s41596-020-0292-x>.
7. Jin S, Guerrero-Juarez CF, Zhang L. et al. Inference and analysis of cell-cell communication using cellchat. *Nat Commun* 2021;**12**:1–13. <https://doi.org/10.1038/s41467-021-21246-9>.
8. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods* 2020;**17**:159–62. <https://doi.org/10.1038/s41592-019-0667-5>.
9. Yuxuan H, Peng T, Gao L. et al. CytoTalk: de novo construction of signal transduction networks using single-cell transcriptomic data. *Science. Advances* 2021;**7**:eabf1356.
10. Hou R, Denisenko E, Ong HT. et al. Predicting cell-to-cell communication networks using NATMI. *Nat Commun* 2020;**11**:5011. <https://doi.org/10.1038/s41467-020-18873-z>.
11. Pham D, Tan X, Jun X. et al. stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. *BioRxiv* 2020;**5**:2020–05.
12. Garcia-Alonso L, Handfield L-F, Roberts K. et al. Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. *Nat Genet* 2021;**53**:1698–711. <https://doi.org/10.1038/s41588-021-00972-2>.
13. Wilk AJ, Shalek AK, Holmes S. et al. Comparative analysis of cell-cell communication at single-cell resolution. *Nat Biotechnol* 2023;**42**:1–14.
14. Fischer DS, Schaar AC, Theis FJ. Modeling intercellular communication in tissues using spatial graphs of cells. *Nat Biotechnol* 2023;**41**:332–6. <https://doi.org/10.1038/s41587-022-01467-z>.
15. Yuan Y, Bar-Joseph Z. GCNG: graph convolutional networks for inferring gene interaction from spatial transcriptomics data. *Genome Biol* 2020;**21**:1–16.
16. Vaswani A, Shazeer N, Parmar N. et al. Attention is all you need. *Adv Neural Inf Process Syst* 2017;**30**:1–13.
17. Chen G, Liu Z-P. Graph attention network for link prediction of gene regulations from single-cell RNA-sequencing data. *Bioinformatics* 2022;**38**:4522–9. <https://doi.org/10.1093/bioinformatics/btac559>.
18. Li H, Ma T, Hao M. et al. Decoding functional cell-cell communication events by multi-view graph learning on spatial transcriptomics. *Brief Bioinform* 2023;**24**:6. <https://doi.org/10.1093/bib/bbad359>.
19. Bi K, He MX, Bakouny Z. et al. Tumor and immune reprogramming during immunotherapy in advanced renal cell carcinoma. *Cancer Cell* 2021;**39**:649–661.e5. <https://doi.org/10.1016/j.ccell.2021.02.015>.
20. Dimitrov D, Türei D, Garrido-Rodriguez M. et al. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-seq data. *Nat Commun* 2022;**13**:3224. <https://doi.org/10.1038/s41467-022-30755-0>.
21. Liu Z, Sun D, Wang C. Evaluation of cell-cell interaction methods by integrating single-cell RNA sequencing data with spatial information. *Genome Biol* 2022;**23**:1–38. <https://doi.org/10.1186/s13059-022-02783-y>.
22. Noël F, Massenet-Regad L, Carmi-Levy I. et al. Dissection of intercellular communication using the transcriptome-based framework icellnet. *Nat Commun* 2021;**12**:1089. <https://doi.org/10.1038/s41467-021-21244-x>.
23. Balluffi RW, Allen SM, Craig W. et al. *Kinetics of Materials*. John Wiley & Sons; 2005:05. <https://doi.org/10.1002/0471749311>.
24. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* 2018;**19**:1–5.
25. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv* 2016;16:1609. 02907.
26. Yuan H, Yan M, Zhang G. et al. CancerSEA: a cancer single-cell state atlas. *Nucleic Acids Res* 2019;**47**:D900–8. <https://doi.org/10.1093/nar/gky939>.
27. Motzer RJ, Robbins PB, Powles T. et al. Avelumab plus axitinib versus sunitinib in advanced renal cell carcinoma: biomarker analysis of the phase 3 JAVELIN renal 101 trial. *Nat Med* 2020;**26**:1733–41. <https://doi.org/10.1038/s41591-020-1044-8>.
28. Braun DA, Hou Y, Bakouny Z. et al. Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma. *Nat Med* 2020;**26**:909–18. <https://doi.org/10.1038/s41591-020-0839-y>.
29. Jeong S, Park S-H. Co-stimulatory receptors in cancers and their implications for cancer immunotherapy. *Immune Netw* 2020;**20**:1–15. <https://doi.org/10.4110/in.2020.20.e3>.
30. Zhang W, Zhang Q, Yang N. et al. Crosstalk between IL-15R $\alpha$ +tumor-associated macrophages and breast cancer cells reduces CD8+T cell recruitment. *Cancer Commun* 2022;**42**:536–57. <https://doi.org/10.1002/cac2.12311>.
31. Tang Y, Kwiatkowski DJ, Henske EP. Midkine expression by stem-like tumor cells drives persistence to mtor inhibition

- and an immune-suppressive microenvironment. *Nat Commun* 2022;**13**:5018. <https://doi.org/10.1038/s41467-022-32673-7>.
32. Puig-Kröger A, Sierra-Filardi E, Domínguez-Soto A. et al. Folate receptor  $\beta$  is expressed by tumor-associated macrophages and constitutes a marker for M2 anti-inflammatory/regulatory macrophages. *Cancer Res* 2009;**69**:9395–403. <https://doi.org/10.1158/0008-5472.CAN-09-2050>.
  33. Zheng P, Luo Q, Wang W. et al. Tumor-associated macrophages-derived exosomes promote the migration of gastric cancer cells by transfer of functional apolipoprotein E. *Cell Death Dis* 2018;**9**:434. <https://doi.org/10.1038/s41419-018-0465-5>.
  34. Hao J, Yan F, Zhang Y. et al. Expression of adipocyte/macrophage fatty acid-binding protein in tumor-associated macrophages promotes breast cancer progression. *Cancer Res* 2018;**78**:2343–55. <https://doi.org/10.1158/0008-5472.CAN-17-2465>.
  35. Pittet MJ, Michielin O, Migliorini D. Clinical relevance of tumour-associated macrophages. *Nat Rev. Clin Oncol* 2022;**19**:402–21. <https://doi.org/10.1038/s41571-022-00620-6>.
  36. Obradovic A, Chowdhury N, Haake SM. et al. Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages. *Cell* 2021;**184**:2988–3005.e16. <https://doi.org/10.1016/j.cell.2021.04.038>.
  37. Hou Y, Wei D, Zhang Z. et al. Fabp5 controls macrophage alternative activation and allergic asthma by selectively programming long-chain unsaturated fatty acid metabolism. *Cell Rep* 2022;**41**:111668. <https://doi.org/10.1016/j.celrep.2022.111668>.
  38. Lopes-Bastos BM, Jiang WG, Cai J. Tumour–endothelial cell communications: important and indispensable mediators of tumour angiogenesis. *Anticancer Res* 2016;**36**:1119–26.
  39. Hills CE, Eleftherios Siamantouras SW, Smith PC. et al. TGF $\beta$  modulates cell-to-cell communication in early epithelial-to-mesenchymal transition. *Diabetologia* 2012;**55**:812–24. <https://doi.org/10.1007/s00125-011-2409-9>.
  40. Calvo F, Sahai E. Cell communication networks in cancer invasion. *Curr Opin Cell Biol* 2011;**23**:621–9. <https://doi.org/10.1016/j.ceb.2011.04.010>.
  41. Claesson-Welsh L, Welsh M. Vegfa and tumour angiogenesis. *J Intern Med* 2013;**273**:114–27. <https://doi.org/10.1111/joim.12019>.
  42. Wolf NK, Kissiov DU, Raulet DH. Roles of natural killer cells in immunity to cancer, and applications to immunotherapy. *Nat Rev Immunol* 2023;**23**:90–105. <https://doi.org/10.1038/s41577-022-00732-1>.
  43. Bald T, Krummel MF, Smyth MJ. et al. The NK cell–cancer cycle: Advances and new challenges in NK cell–based immunotherapies. *Nat Immunol* 2020;**21**:835–47. <https://doi.org/10.1038/s41590-020-0728-z>.
  44. Abel AM, Yang C, Thakar MS. et al. Natural killer cells: development, maturation, and clinical utilization. *Front Immunol* 2018;**9**:1869. <https://doi.org/10.3389/fimmu.2018.01869>.
  45. Lambert AW, Pattabiraman DR, Weinberg RA. Emerging biological principles of metastasis. *Cell* 2017;**168**:670–91. <https://doi.org/10.1016/j.cell.2016.11.037>.
  46. Wei Q, Qian Y, Jun Y. et al. Metabolic rewiring in the promotion of cancer metastasis: mechanisms and therapeutic implications. *Oncogene* 2020;**39**:6139–56. <https://doi.org/10.1038/s41388-020-01432-7>.
  47. Tian H, Lian R, Li Y. et al. AKT-induced lncRNA VAL promotes EMT-independent metastasis through diminishing TRIM16-dependent vimentin degradation. *Nat Commun* 2020;**11**:5127. <https://doi.org/10.1038/s41467-020-18929-0>.
  48. Maishi N, Hida K. Tumor endothelial cells accelerate tumor metastasis. *Cancer Sci* 2017;**108**:1921–6. <https://doi.org/10.1111/cas.13336>.
  49. Nakamura K, Smyth MJ. Immunoediting of cancer metastasis by NK cells. *Nature cancer* 2020;**1**:670–1. <https://doi.org/10.1038/s43018-020-0081-z>.
  50. Liu X, Song J, Zhang H. et al. Immune checkpoint HLA-E: CD94-NKG2A mediates evasion of circulating tumor cells from NK cell surveillance. *Cancer Cell* 2023;**41**:272–287.e9. <https://doi.org/10.1016/j.ccell.2023.01.001>.
  51. Lin Q, Rong L, Jia X. et al. IFN- $\gamma$ -dependent NK cell activation is essential to metastasis suppression by engineered salmonella. *Nat Commun* 2021;**12**:2537. <https://doi.org/10.1038/s41467-021-22755-3>.
  52. Lo HC, Zhan X, Kim IS. et al. Resistance to natural killer cell immunosurveillance confers a selective advantage to polyclonal metastasis. *Nat Cancer* 2020;**1**:709–22.
  53. Dallavalasa S, Beeraka NM, Basavaraju CG. et al. The role of tumor associated macrophages (TAMS) in cancer progression, chemoresistance, angiogenesis and metastasis-current status. *Curr Med Chem* 2021;**28**:8203–36. <https://doi.org/10.2174/0929867328666210720143721>.
  54. Chen Y, Song Y, Wei D. et al. Tumor-associated macrophages: an accomplice in solid tumor progression. *J Biomed Sci* 2019;**26**:1–13.
  55. Borriello L, Coste A, Traub B. et al. Primary tumor associated macrophages activate programs of invasion and dormancy in disseminating tumor cells. *Nat Commun* 2022;**13**:626. <https://doi.org/10.1038/s41467-022-28076-3>.
  56. Ren L, Yi J, Yang Y. et al. Systematic pan-cancer analysis identifies APOC1 as an immunological biomarker which regulates macrophage polarization and promotes tumor metastasis. *Pharmacol Res* 2022;**183**:106376. <https://doi.org/10.1016/j.phrs.2022.106376>.
  57. An L, Liu Y. ZNF460 mediates epithelial-mesenchymal transition to promote gastric cancer progression by transactivating APOC1 expression. *Exp Cell Res* 2023;**422**:113452. <https://doi.org/10.1016/j.yexcr.2022.113452>.
  58. Li Y-L, Lin-wen W, Zeng L-H. et al. APOC1 promotes the metastasis of clear cell renal cell carcinoma via activation of STAT3. *Oncogene* 2020;**39**:6203–17. <https://doi.org/10.1038/s41388-020-01428-3>.
  59. Sun Y-F, Liang W, Liu S-P. et al. Dissecting spatial heterogeneity and the immune-evasion mechanism of ctcs by single-cell RNA-seq in hepatocellular carcinoma. *Nat Commun* 2021;**12**:4091. <https://doi.org/10.1038/s41467-021-24386-0>.
  60. Gate D, Saligrama N, Leventhal O. et al. Clonally expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer's disease. *Nature* 2020;**577**:399–404. <https://doi.org/10.1038/s41586-019-1895-7>.
  61. Pipkin ME, Sacks JA, Cruz-Guilloty F. et al. Interleukin-2 and inflammation induce distinct transcriptional programs that promote the differentiation of effector cytolytic T cells. *Immunity* 2010;**32**:79–90. <https://doi.org/10.1016/j.immuni.2009.11.012>.