# CytoTalk

**Version 1.0**
**Mar 4th, 2020**

## Overview

Cell-cell communication in a tissue microenvironment is mediated by signal transduction pathways. Single-cell technology has opened the door for studying signal transduction at much higher resolution in a complex tissue. Currently, there is a lack of analytical methods to infer signal transduction pathways based on single-cell omics data. Here we introduce a computational method, CytoTalk, to construct signal transduction networks using single-cell RNA-Seq data. The method first constructs intracellular and intercellular gene-gene interaction networks using an information-theoretic measure between two cell types. Candidate signal transduction pathways in the integrated network are identified using the Prize-Collecting Steiner Forest (PCSF) algorithm. CytoTalk is implemented using MATLAB (version > R2017a), R (version > 3.5.0) and Python (version 2.7).

## I. Input files

**(1)** CytoTalk requires a comma-delimited ".csv" file containing scRNA-Seq data for each cell type under study. The files should be named as "scRNAseq_CellTypeA.csv", "scRNAseq_CellTypeB.csv", "scRNAseq_CellTypeC.csv", etc. Each file contains the **log2-transformed normalized scRNA-Seq data** for a cell type with rows as genes (GENE SYMBOL) and columns as cells. Examples are in the folder /ExampleInput/ folder.

Csv files for all cell types should be **copied into the /CytoTalk/ folder (NOT the /ExampleInput/ folder)** since all of them are needed for computing cell-type-specificity of gene expression in the CytoTalk algorithm.

!!!Note that the /CytoTalk/ folder can only be used once for identifying the communication gene module (signaling networks) between cell type A and cell type B. If you also want to identify the module between cell types C and D, you will need to re-name scRNA-Seq data of cell types C and D as A and B, respectively, and re-name original cell types A and B as C and D, respectively. Then, copy a cleaned /CytoTalk/ folder to your working directory.

**(2)** CytoTalk also requires a txt file indicating the species from which the scRNA-Seq data are generated. Currently, "Human" and "Mouse" are supported. An example is in the folder /ExampleInput/ folder. This "Species.txt" should be also **copied into the /CytoTalk/ folder (NOT the /ExampleInput/ folder).**

**II. Required R and Python packages and system environment variables**

1. The following four R packages should be installed (R version > 3.5.0 is recommended).

1) "entropy": https://cran.r-project.org/web/packages/entropy/index.html

2) "infotheo": https://cran.r-project.org/web/packages/infotheo/index.html

3) "doParallel": https://cran.r-project.org/web/packages/doParallel/index.html

This package is used for parallel computation of mutual information for all gene pairs. Before running CytoTalk, change the number of logical cores available in Line 13 of the R scripts "comp_MIcoexp_TypA_WinPara.R" and "comp_MIcoexp_TypB_WinPara.R". The default is 6 logical cores.

4) "parmigene": https://cran.rstudio.com/web/packages/parmigene/index.html

This package is used for parallel computation of indirect edge-filtered gene networks. Before running CytoTalk, set system environment variable as following:
**export OMP_NUM_THREADS=*n***, where *n* is the number of logical cores available.

2. The following Python package should be installed.

"pcsf_fast": https://github.com/fraenkel-lab/pcst_fast

This package is used for fast identification of a rooted Prize-collecting Steiner tree in a network. Before running CytoTalk, set the system environment variable as following:
**export PYTHONPATH=$PYTHONPATH:/your installed pcsf_fast folder/**


**III. Running CytoTalk**

Copy the "CytoTalk/" directory under the "CytoTalk_package_xxxx" folder to your working directory and execute the following two steps:

1) **bash CommunModule_TypATypB_Step1.sh**

This step needs to call MATLAB function. So you may need to **specify the absolute path of the executable MATLAB program** in the Line 14 of this bash script. This step may take up to 6 hours because computing mutual information for all gene pairs

is time consuming.

2) **bash CommunModule_TypATypB_Step2.sh**

This step needs to call MATLAB function. So you may need to **specify the absolute path of the executable MATLAB program** in the Line 13 and Line 23 of this bash script.

## IV. CytoTalk output

The output folder, "/CytoTalk/IllustratePCSF/**"**, contains a network file "PCSF_edgeSym.sif" and the following six attribute files that are ready for import into Cytoscape for visualization and further analysis of the signaling network identified between cell type A and cell type B.

Two edge attribute files:

1) "PCSF_edgeCellType.txt"
2) "PCSF_edgeCost.txt"

Four node attribute files:

1) "PCSF_geneCellType.txt"
2) "PCSF_geneExp.txt"
3) "PCSF_genePrize.txt"
4) "PCSF_geneRealName.txt"

**Contact:**

Kai Tan, tank1@email.chop.edu
Yuxuan Hu, yuxuan_hu_xd@163.com