**scPAIR-seq**

scPAIR-seq is a computational and experimental single cell framework for analyzing PAthogen-specific Immune Responses. It combines scRNA-seq of infected host cells with multiplex-tagged mutant bacterial library, to identify mutant-specific changes in host transcriptome.

For more information: <https://doi.org/10.1101/2022.03.06.483158>

System requirement: MATALB and R

**The pipeline contains two parts**:

1) Identifying the identity of the invading bacteria using the bacterial Mutant Identifier Barcode (MIB), processed from the single cell bacterial library.

2) Identify changes in host transcriptome due to mutant identity (MAESTRO analysis), using the single cell host library.

Examples for input files can be found in: input\_data/

**PART 1 – single cell bacterial library**:

*install\_packages.R*: contains the relevant packages for this part.

*generate\_MIB\_matrix.R*: gets as input R1 and R2 fastq files (e.g., ex\_R1.fastq and ex\_R2.fastq), and generates a MIB count matrix (i.e. number of reads classified to each bacterial mutant in each cell).

*classify\_infected\_single\_cells.R*: gets as input the MIB count matrix from the previous step, or a normalized MIB count matrix (e.g., ex\_norm\_MIB\_matrix.txt), and outputs the identity of the invading mutant for each cell, based on the distribution of each MIB across all single cells (using the MULTI-seq package <https://github.com/chris-mcginnis-ucsf/MULTI-seq>, McGinnis et al., 2019).

**PART 2 – single cell host library**:

*MAESTRO\_functions.R*: contains all functions required for *calc\_MAESTRO\_scores.R*.

*calc\_MAESTRO\_scores.R*: gets as input the host single cell RNA-seq UMI counts matrix with the identity of the infecting bacteria for each cell, as generated by *classify\_infected\_single\_cells.R* (e.g., sce\_norm.RDS). Calculates the MAESTRO scores for each gene in each mutant, relative to WT infected cells. Generates a random model from all host cells for which we could not identify the identity of the invading bacteria. Outputs a table with the MAESTRO scores of the real data (maestro\_scores\_real\_data.txt) and the random data (maestro\_scores\_rand.txt).

*MAESTRO\_analysis.m*: gets as input the tables of the real and random data that were generated in *calc\_MAESTRO\_scores.R* and calculates p-value and q-value for each mutant based on MAESTRO scores distribution. Outputs a table with statistics about the significant mutants.

*extract\_DEG.m*: MATLAB function which gets as input the name of the mutant and outputs its DEGs separated to up and down-regulated relative to WT infected cells. The function uses as input file the real MAESTRO scores file generated in *calc\_MAESTRO\_scores.R*.

**To reproduce the data presented at the manuscript:**1) run the script *classify\_infected\_single\_cells.R* on the normalized MIB count matrix ex\_norm\_MIB\_matrix.txt

2) run the script *MAESTRO\_analysis.m* with the output files from *calc\_MAESTRO\_scores.R*: maestro\_scores\_real\_data.txt and maestro\_scores\_rand.txt

3) run the script *extract\_DEG.m* to extract the DEGs of each mutant.