

SCODA pipeline

for single-cell transcriptomics data analysis

MLBI Lab

SCODA?

- ❑ Web-based, Fully-automated, all-in-one computing service for **S**ingle-**C**ell (transcript)**O**mics **D**ata **A**nalysis (single-cell RNA-seq)
- ❑ It is useful for
 - Tissue/tumor microenvironment study to elucidate underlying biological mechanism
 - Discovery of therapeutic targets in many diseases & cancers

How SCODA works?

Web-based, automated analysis pipeline: SCODA

Single-cell RNA-seq raw data

MLBI LAB SCODA pipeline (Single-Cell Omics Data Analysis Pipeline)

SCODA is a web-based, fully automated, all-in-one analysis pipeline for single-cell RNA-seq data.

Discover the Hidden Secrets in Your Single-Cell RNA-seq Data with Our Cutting-Edge Analysis Pipeline

Simply upload your single-cell RNA-seq datasets in a compressed file, and you will get the results in an AnViz formatted HTML file which can be loaded in Python or R to visualize and interact with your analysis results. The results include:

1. Cell-type annotation in 3-level taxonomy, i.e., major-type, minor-type and sub-type;
2. CNV estimation and tumor cell annotation using inferCNVpy (RISC 3-level);
3. Cell-cell interaction analysis per condition and per cell-type using CellPhoneDB (RISC 3-level);
4. Differentially expressed gene (DEG) analysis results per condition and per cell-type;
5. Gene set analysis (GSEA) results per condition and per cell-type using GSEApy (RISC 3-level).

Once you download the result file, you can load it in the example Jupyter notebook we have hidden in your data. With optional configuration setting, you can tailor the analysis workflow for your specific data. Download/Check pipeline description here.

Quick trial steps:

1. Download an example dataset here.
2. Upload it through the form below.
3. Once it is done, download the result (xxx.tar.gz).
4. Open Jupyter notebook in Google Colab by clicking here.
5. Upload the xxx.tar.gz file and follow the instruction there.

Mandatory Input:

The following 3 files:

- 1. Compressed H5AD file
- 2. Compressed cell-type annotation file
- 3. Compressed cell-type annotation file

Optional Input:

- 1. markers_db.tsv
- 2. cellphonedb.zip
- 3. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Optional Input:

- 1. cellphonedb.zip
- 2. pathways.gmt

Functional Analysis

Cell-type annotation
CNV estimation
Identifying tumor cells
Cell-cell interaction
Diff. expressed gene
Gene set analysis
Marker discovery

Download result file with functional annotations

Client's own data mining using SCODA-viz package

Jupyter Notebook for client's own data mining

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

SCODA-viz package

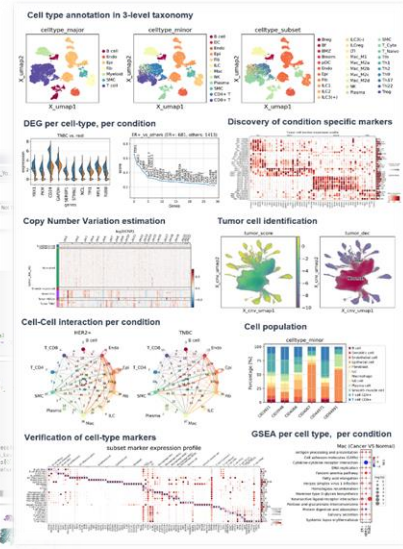
SCODA-viz package

SCODA-viz package

SCODA-viz package

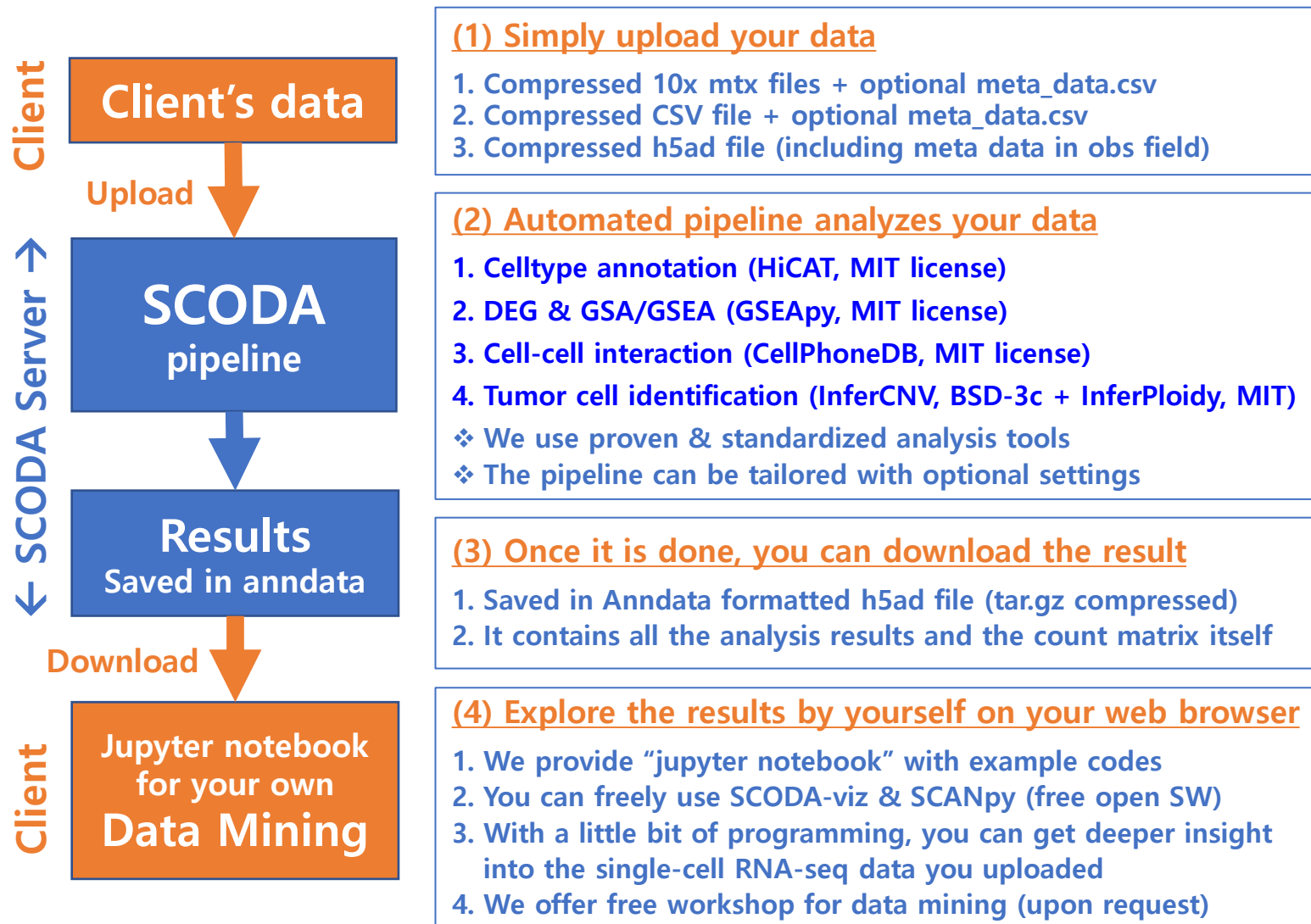
SCODA-viz package

SCODA-viz package

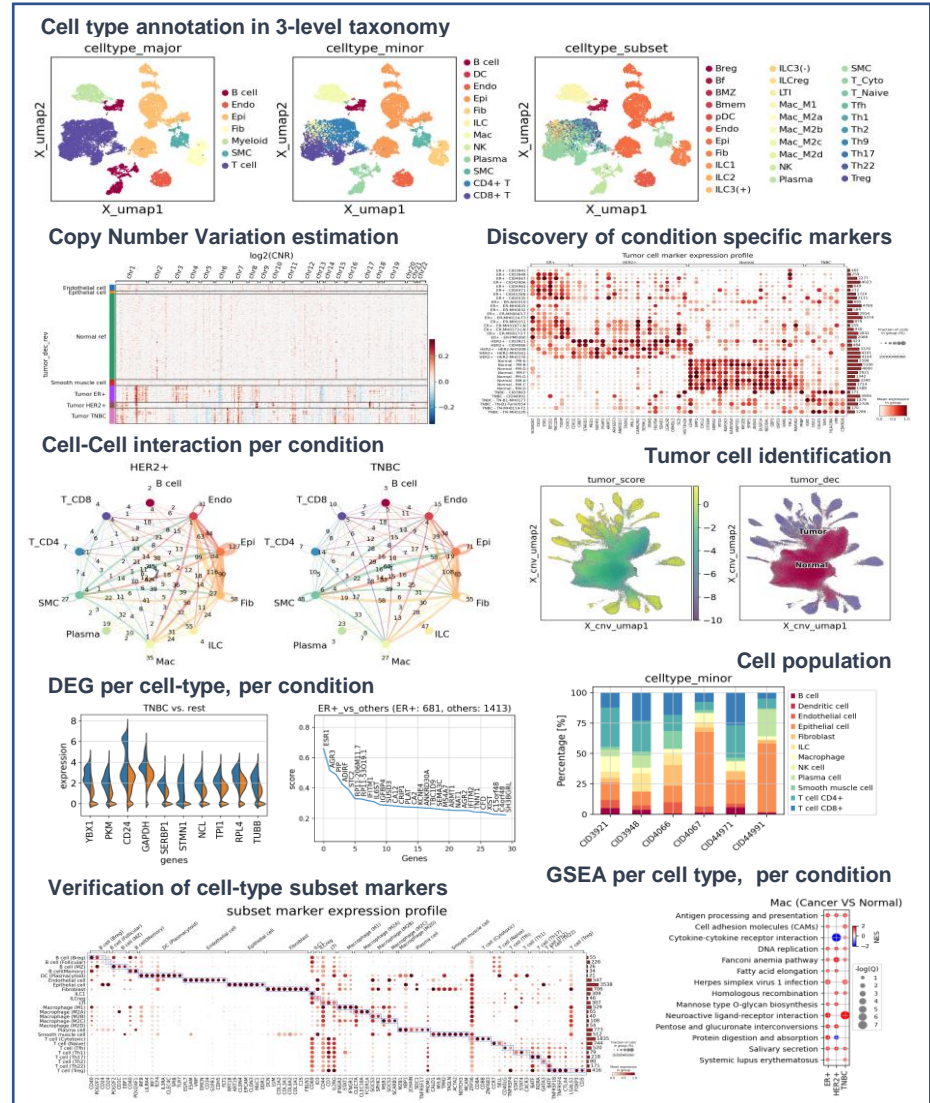
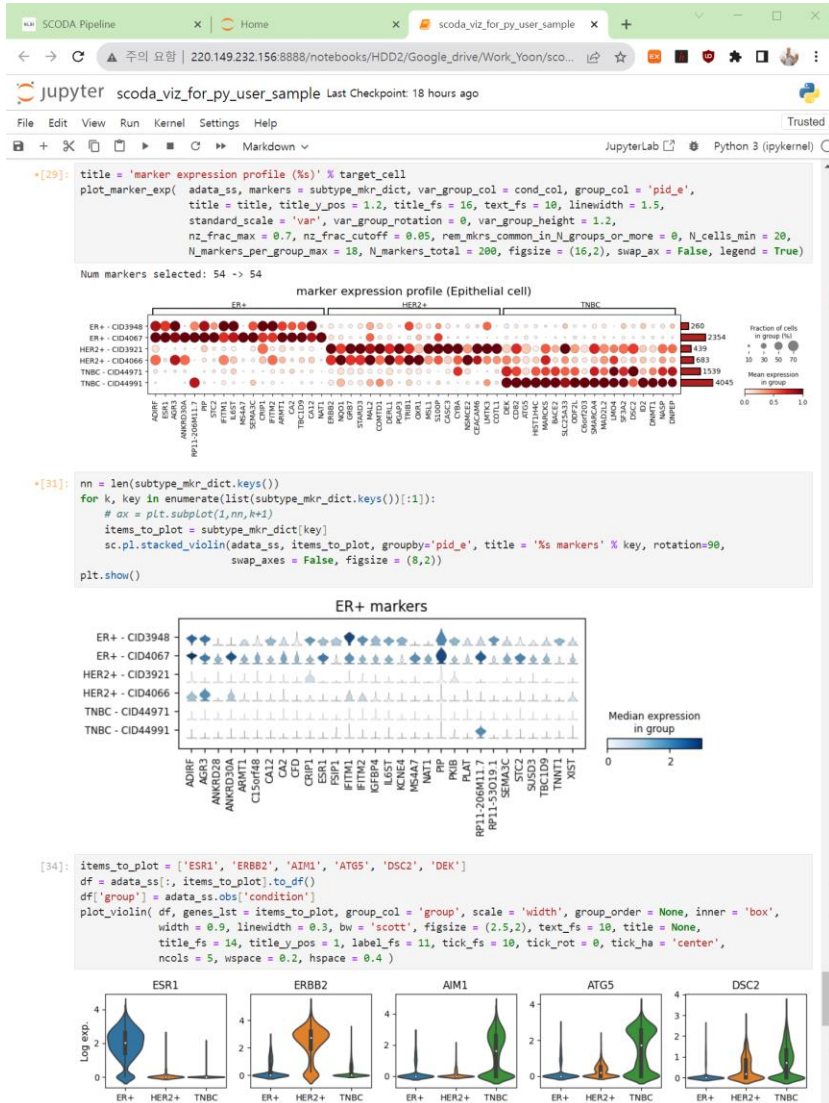


- Therapeutic target discovery
- Biological mechanism
- Design validation experiment

How SCODA works?



Visualization & data mining



Using SCODA

- ❑ SCODA homepage: <https://mlbi-lab.net>
- ❑ MLBI lab company homepage: <https://mlbi-lab.com>

Why SCODA?

Bio/Medical background

+

SCODA

with SCODA-viz tool

+

Little bit of programming skill
(Free Training workshop available)

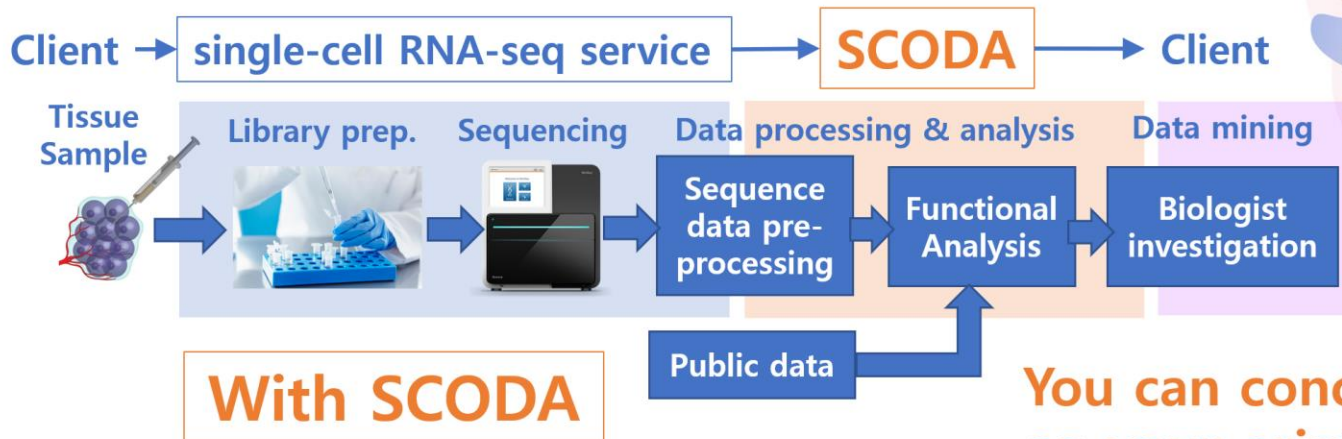
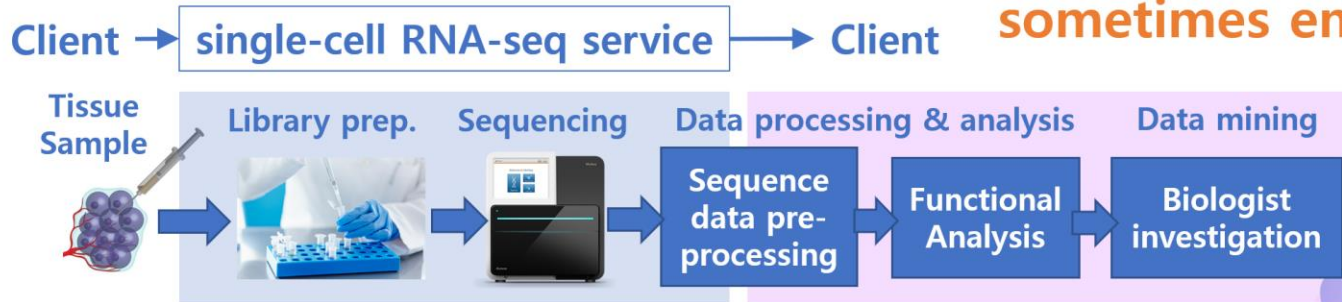
||

Single-cell RNA-seq
Data analysis expert

Why SCODA?

Without SCODA

You must do something
that is not your original job
It is time-consuming &
sometimes embarrassing



You can concentrate more
on your original jobs!!

Related papers

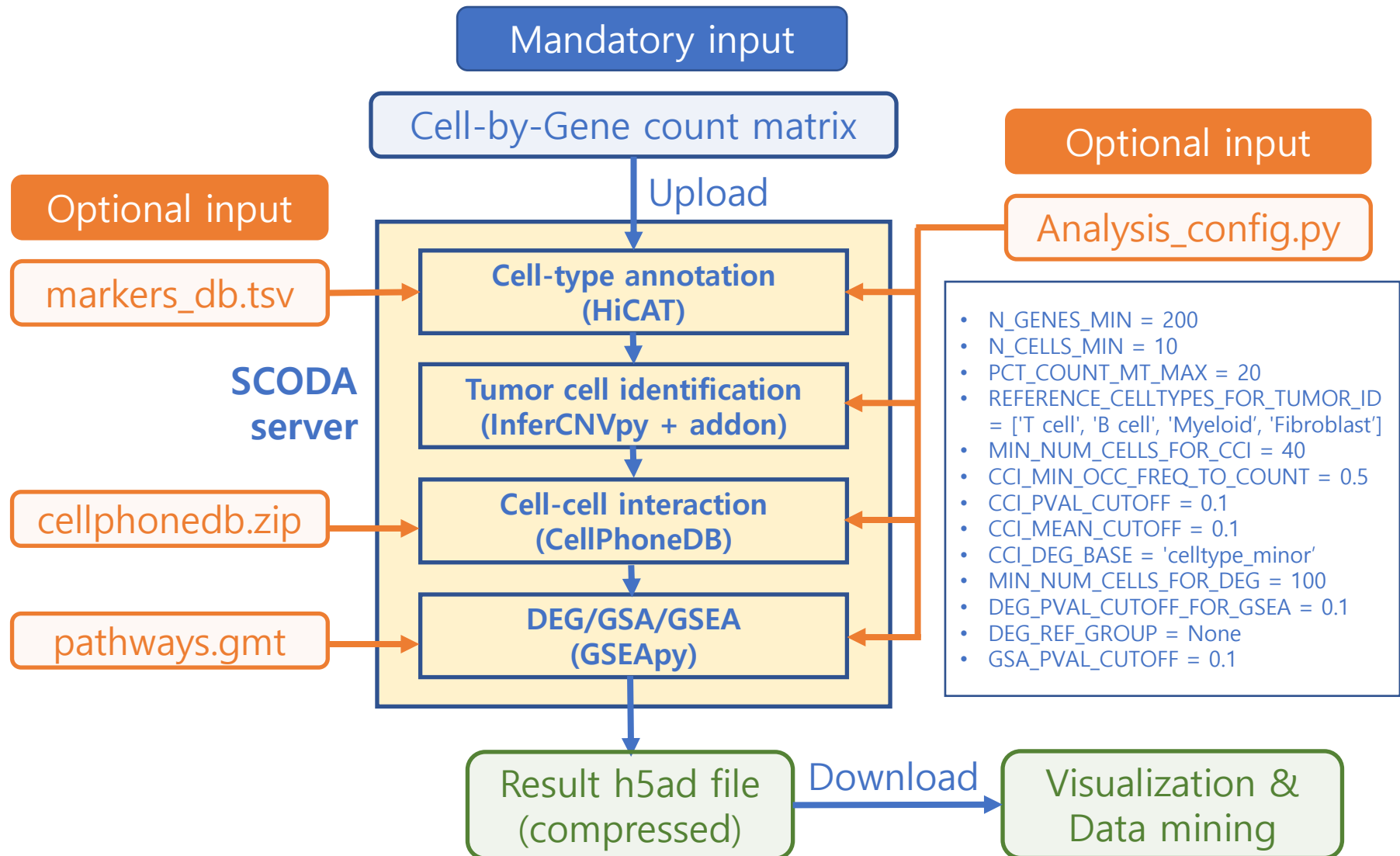
□ HiCAT

- Hierarchical cell-type identifier accurately distinguishes immune-cell subtypes enabling precise profiling of tissue microenvironment with single-cell RNA-sequencing, **Briefings in bioinformatics, March 2023**

□ Studies using SCODA

- Colon-Targeted eNAMPT-Specific Peptide Systems for Treatment of DSS-Induced Acute and Chronic Colitis in Mouse, **Antioxidants, Nov. 2022**
- Integrative analysis of ulcerative colitis progression using single-cell RNA-seq and microbiome, **Communications Biology, June 2024**
- A Retrospective View on Triple Negative Breast Cancer Microenvironment: Novel Markers, Interactions, and Mechanisms of Tumor-Associated Components using public Single-cell RNA Seq Datasets, **Cancers, May 2024**

SCODA Optional configuration



SCODA?

❑ SCODA is useful for

- In-vivo tissue/tumor micro-environment (TME) study.
- Immune cell profiling in many diseases, e.g., autoimmune disease & cancer
- Discovery of diagnostic/prognostic markers
- Discovery of druggable targets and its biological mechanism around pathological tissue
- Exploring drug response and mechanism of action

❑ But not suitable yet for

- Studies with cell line
- Differentiation study

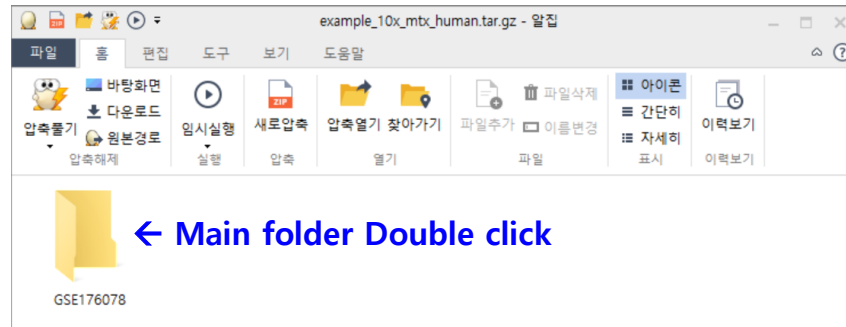
Summary

- ❑ **SCODA utilizes proven open-source software**
 - **HiCAT (MIT license)** for cell type annotation
 - **CellPhoneDB (MIT license)** for inferring cell-cell interaction
 - **InferCNVpy (BSC-3clause)** for CNV estimation
 - **GSEAPy (BSC-3clause)** for gene set enrichment analysis
- ❑ **SCODA-viz package and example jupyter notebook freely available for visualization and data mining**
 - With a little bit of programming skill, you can create any kind of plots you want. (Free training workshop available upon request)
- ❑ **It accelerate your research with single-cell RNA-seq experiment, saving your time and the cost.**
 - Use SCODA first to get insight into the tissue of your interest.
 - Then, plan biological experiment to verify your hypothesis.

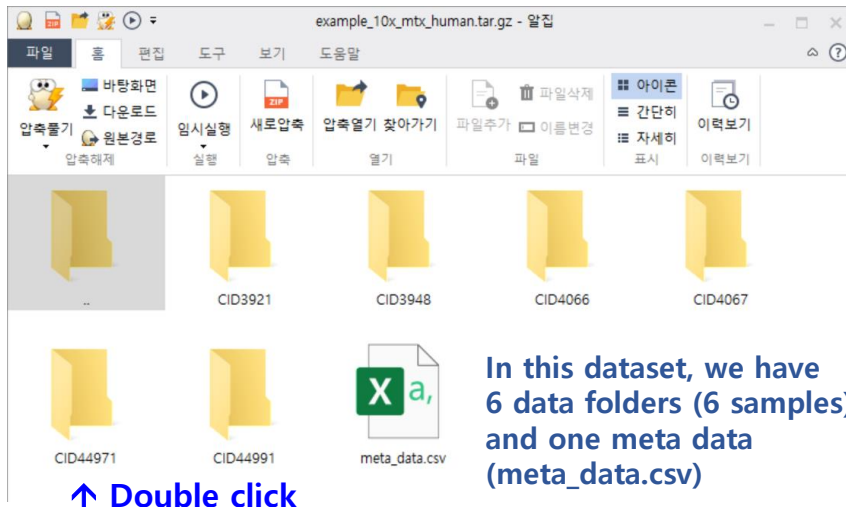
Thank you

Input data formatting (1) 10x_mtx

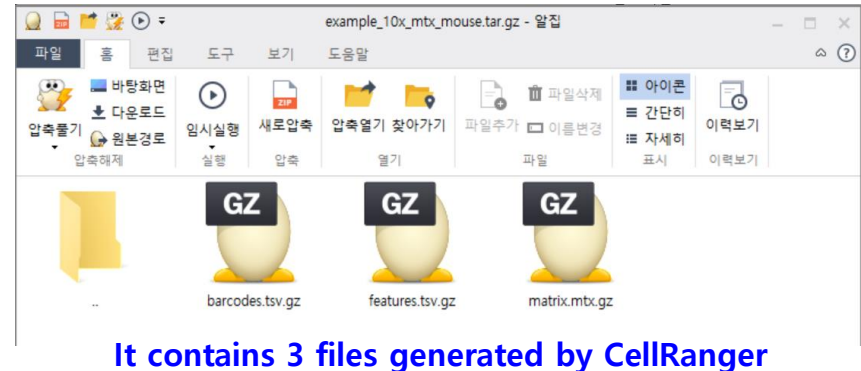
1. Contents of the compressed input file (.zip or .gz)



2. Opening the main folder, we see the following



3. Further opening each data folder, we see



4. Opening the meta_data.csv

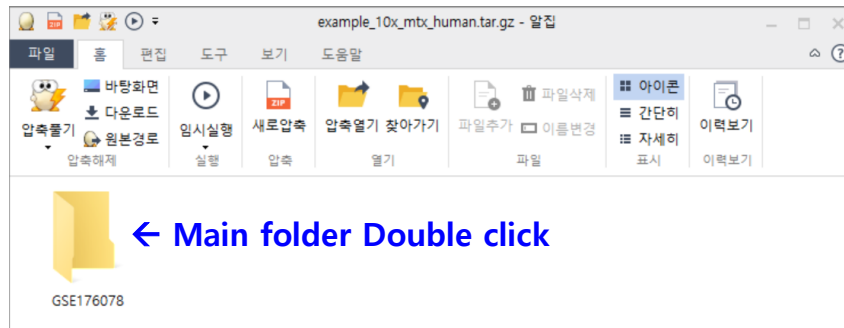
	A	B	C	D
1		sample	condition	
2	CID3921	CID3921	HER2+	
3	CID3948	CID3948	ER+	
4	CID4066	CID4066	HER2+	
5	CID4067	CID4067	ER+	
6	CID44971	CID44971	TNBC	
7	CID44991	CID44991	TNBC	
8				
9				
10				

A: Index column
B: sample name
C: condition

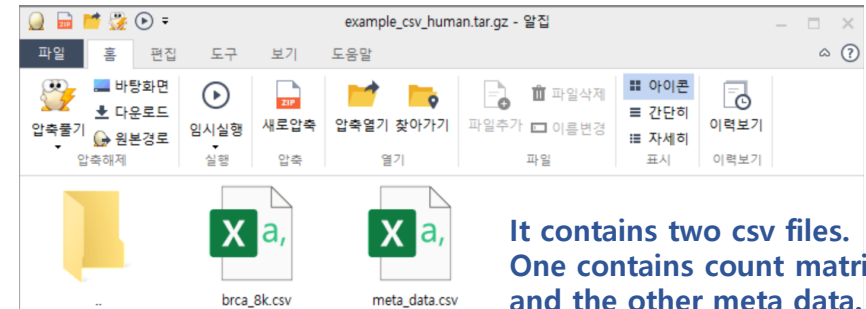
- Each index corresponds to one of the folders (name).
- But we don't require sample name must correspond to a folder name.
- Condition is required to perform DEG, GSEA, cell-cell interaction to compare difference among conditions

Input data formatting (2) csv format

1. Contents of the compressed input file (.zip or .gz)



2. Opening the main folder, we see the following



It contains two csv files.
One contains count matrix
and the other meta data.
(The latter must be named
"meta_data.csv")

3. Data csv file (containing count matrix)

Hugo
symbol

Count
matrix

Cell barcode (cell ID)

4. meta_data.csv 파일

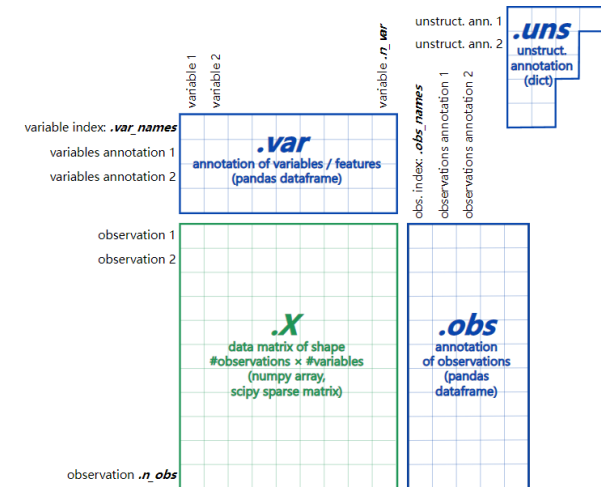
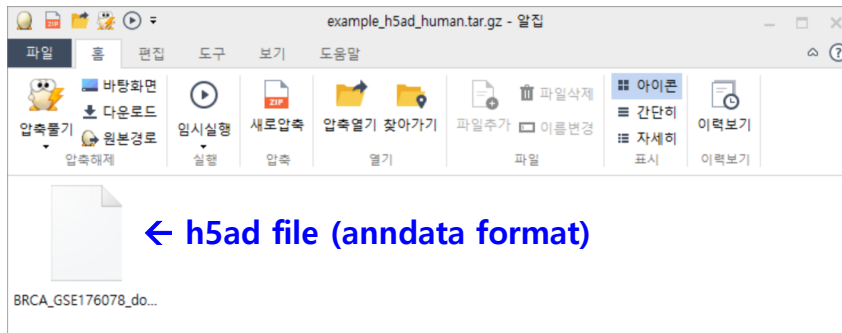
Cell barcode (cell ID)

A: Index column
B: sample
C: condition
D~: optional items

- The indices of the data matrix must have 1-to-1 correspondence to those of meta data matrix.
- Condition is required to perform DEG, GSEA, cell-cell interaction to compare difference among conditions

Input data formatting (3) h5ad format

1. Contents of the compressed input file (.zip or .gz)



2. h5ad file contents

<https://anndata.readthedocs.io/en/latest/>

```
adata_t = sc.read_h5ad(file_h5ad)
adata_t
```

AnnData object with n_obs × n_vars = 12000 × 29733

obs: 'Patient', 'Percent_mito', 'nCount_RNA', 'nFeature_RNA', 'Celltype_Major', 'Celltype_Minor', 'Celltype_Subset', 'subtype', 'gene_module', 'Calls', 'normal_cell_call', 'CNA_value', **'sample', 'condition'**
var: 'gene_ids'

- AnnData contains “sample” and “condition” columns to run DEG/GSEA. DEG/GSEA will not be performed if the obs field does not contain both “sample” and “condition” column.
- If the “sample” column exists in the obs field, cell-cell interaction will be performed per-sample the same as in the above.