

# 웹기반의 단일세포 RNA-seq 데이터 자동분석 서비스 SCODA™를 이용한 종양/조직 미세환경 연구

윤석현

(주) 엠엘비아이랩 (MLBI Lab)

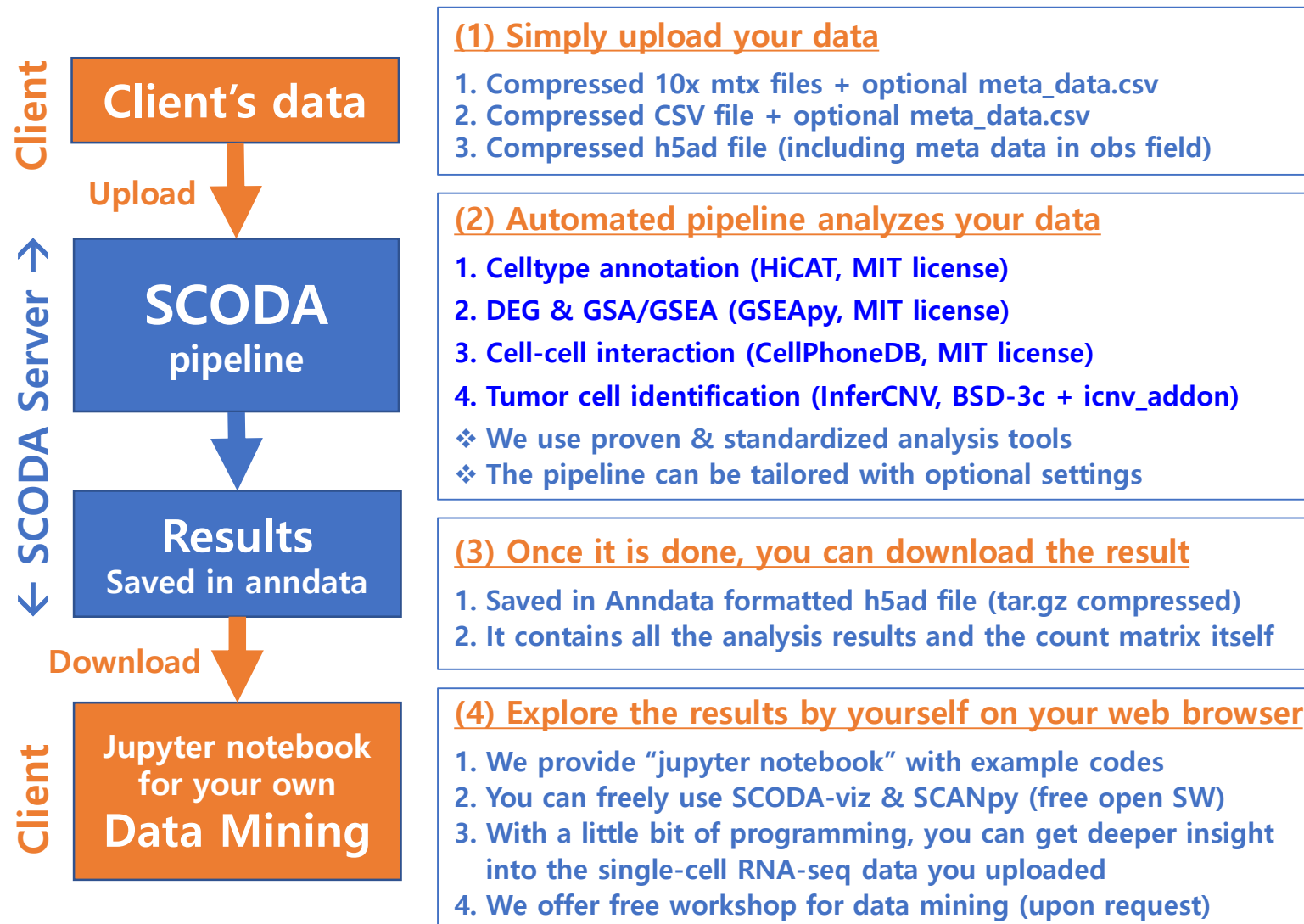
단국대학교 전자전기공학부

단국대학교 대학원 인공지능융합학과

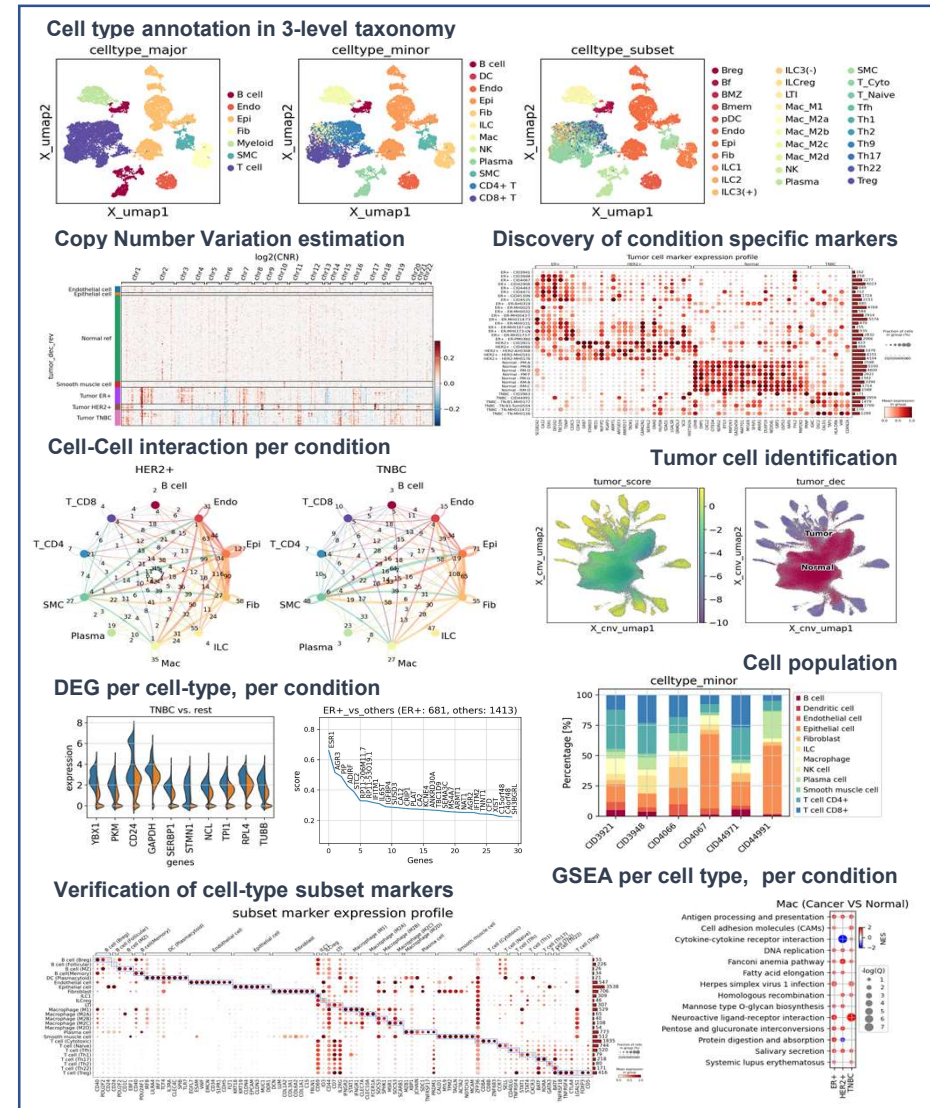
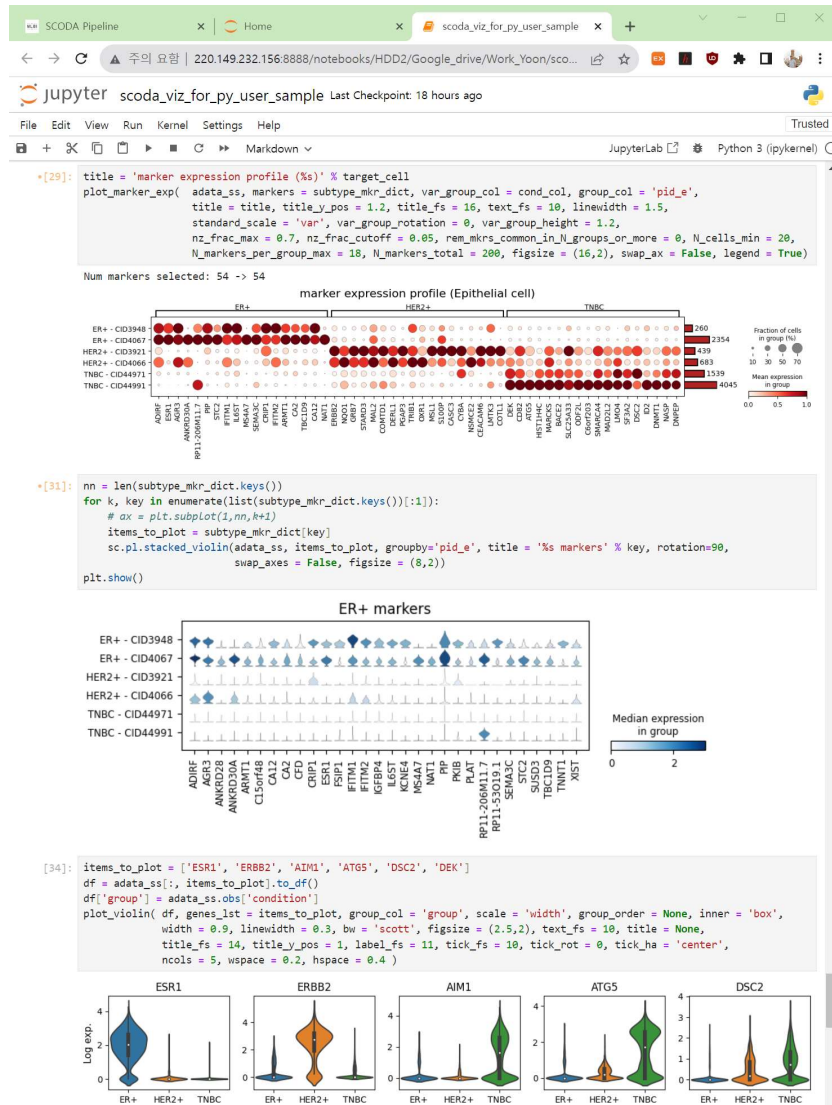
# SCODA?

- ❑ Web-based, Fully-automated, all-in-one computing service for **Single-Cell** (transcript) **O**mic **D**ata **A**nalysis (single-cell RNA-seq)
- ❑ It is useful for
  - In-vivo tissue/tumor micro-environment (TME) study.
  - Immune cell profiling in many diseases, e.g., autoimmune disease & cancer

# 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)

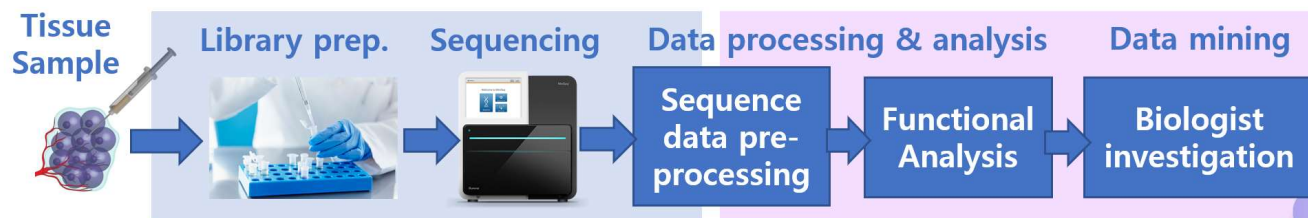
||

Single-cell RNA-seq  
데이터 분석 전문가

# Why SCODA?

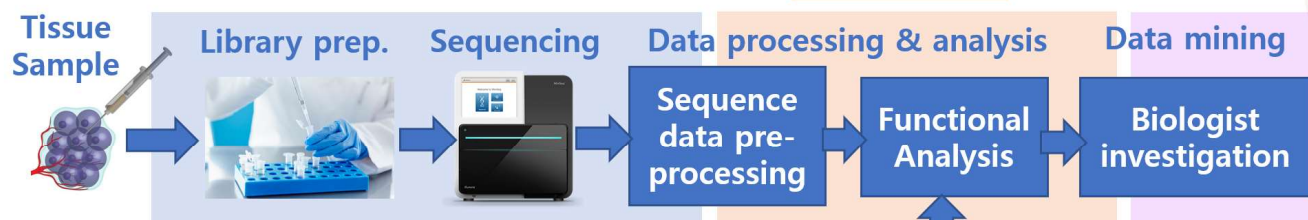
## Without SCODA

Client → single-cell RNA-seq service → Client



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

Client → single-cell RNA-seq service → SCODA → Client



## With SCODA

You can concentrate more  
on your original jobs!!

# Related papers

## □ HiCAT

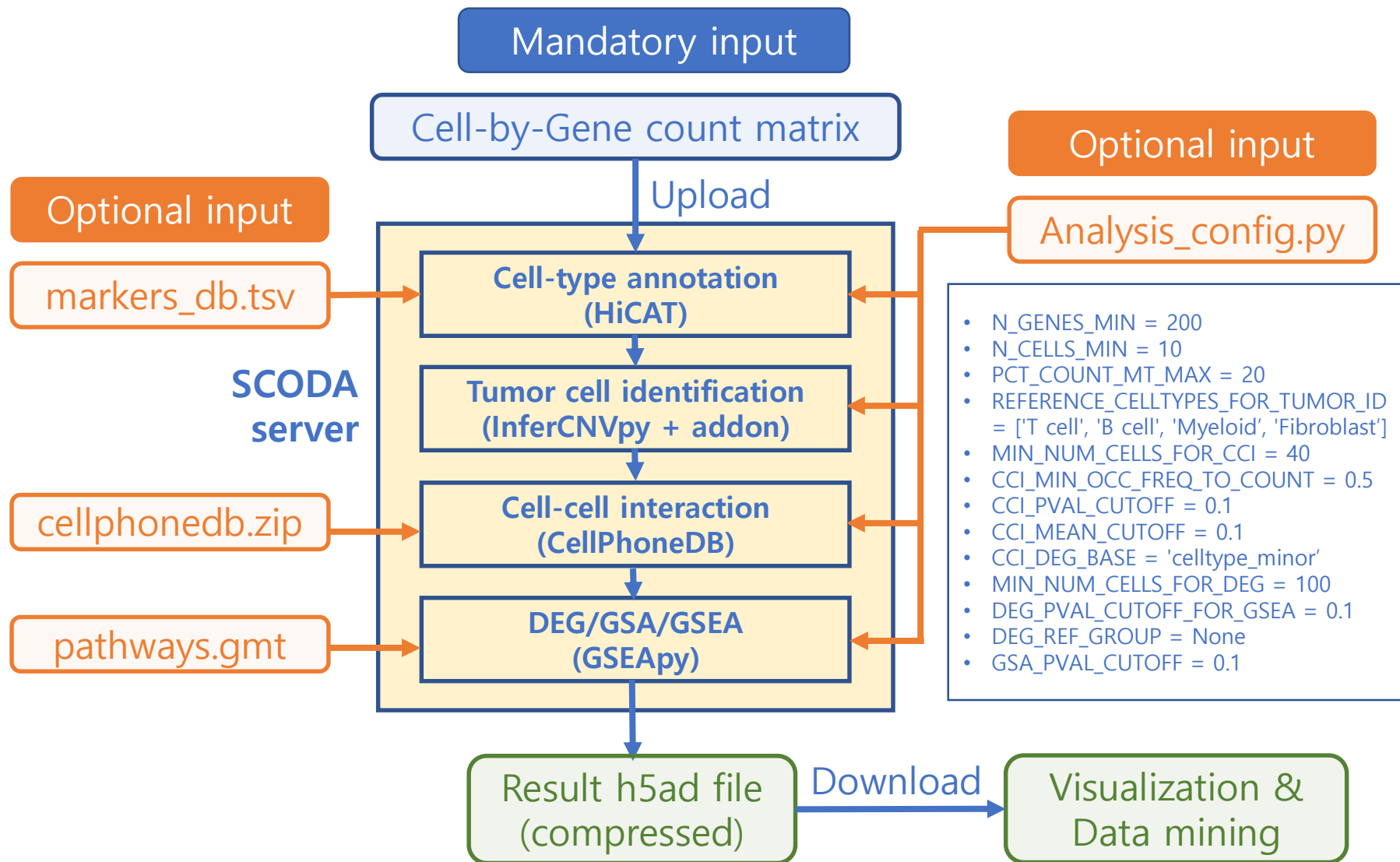
- Hierarchical cell-type identifier accurately distinguishes immune-cell subtypes enabling precise profiling of tissue microenvironment with single-cell RNA-sequencing, **Briefings in bioinformatics, March 2023**

## □ 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, in revision**
- 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, submitted**



# 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.

# Service & Cost

## ❑ Pricing

- 유효 셀 당 ₩20 (unassigned 포함, 2024년 6월 30일 까지)
- 자동 결제 시스템 구축 중 (3월 중 오픈 예정)

## ❑ Our primary concern is client's satisfaction

- 동일 데이터에 대해 (optional configuration 변경하여) 추가 결제 없이 재분석 가능 (6개월 이내 8회)
- 3시간 무료 training workshop (주피터 노트북 사용법, 데이터 시각화 관련 내용)

## ❑ Suggestion & request

- SCODA 데모 페이지에서 먼저 보유하고 계신 데이터를 테스트 해보신 후 (대체적으로 예상되는 결과를 미리 보시고) Full Service 요청하면 좋을 듯 합니다.
- SCODA 데모 분석 결과에 대한 문의/의견/Bug report 등 환영 합니다.
- 가능한 한 의견들 반영하여 당분간 지속적으로 업그레이드 할 예정입니다.

## ❑ Any inquiries?

- 070-7766-5841
- inquiry@mlbi-lab.com

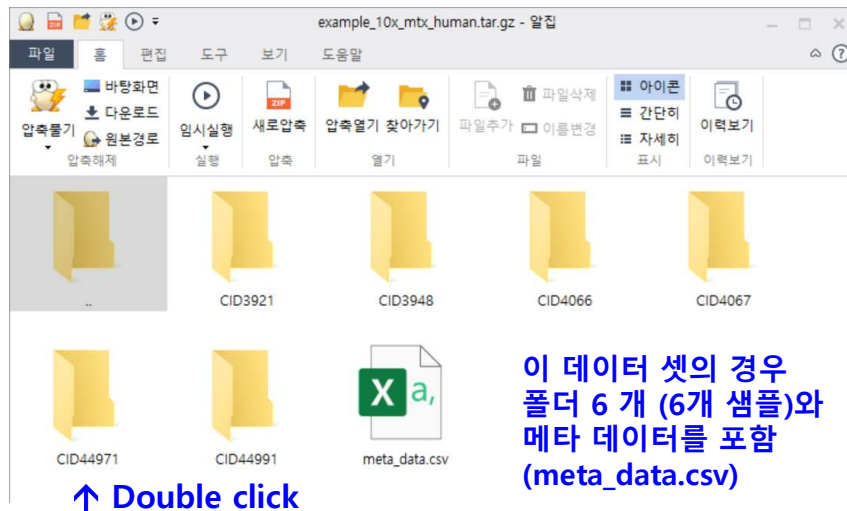
# Thank you

# Input data formatting (1) 10x\_mtx

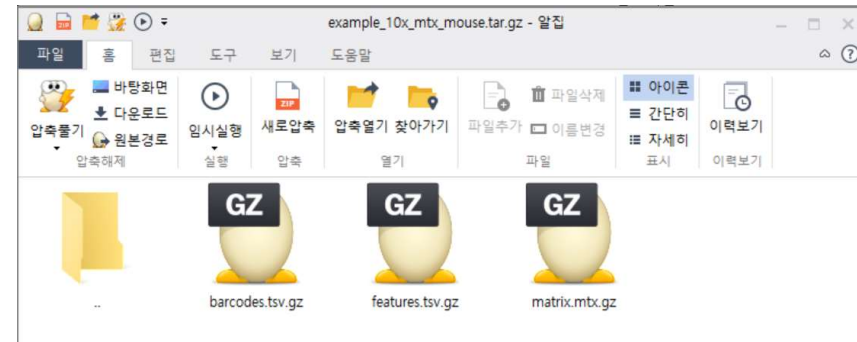
## 1. 압축된 예제 데이터셋을 알집으로 열어 보면



## 2. 메인 폴더를 더블 클릭해서 열면



## 3. 각 데이터 폴더를 더블 클릭해서 열면



CellRanger로 생성된 3개 파일을 포함

## 4. meta\_data.csv 파일을 엑셀로 열어 보면

|    | 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 |          |          |           |   |

1열: 인덱스 열  
2열: sample (name)  
3열: condition

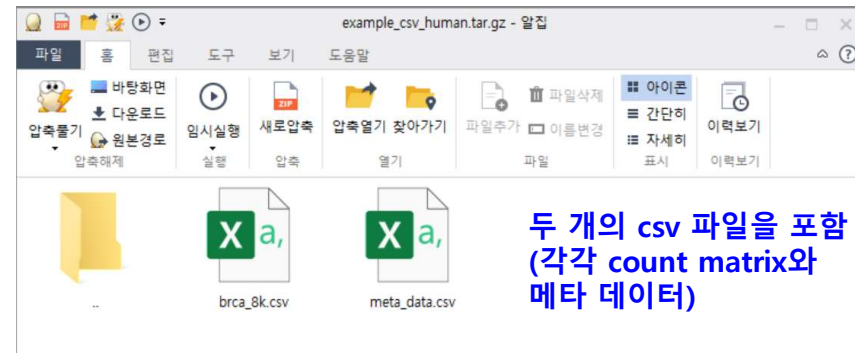
- 인덱스 열의 각 인덱스는 폴더명과 일치해야 함.
- Sample name과 인덱스가 동일할 필요는 없음.
- Condition 열의 조건들을 대상으로 DEG, GSEA, cell-cell interaction 비교가 수행됨

# Input data formatting (2) csv format

## 1. 압축된 예제 데이터셋을 알집으로 열어 보면



## 2. 메인 폴더를 더블 클릭해서 열면



## 3. 데이터 csv 파일

Cell barcode (cell ID)

Count matrix

Hugo symbol

## 4. meta\_data.csv 파일

Cell barcode (cell ID)

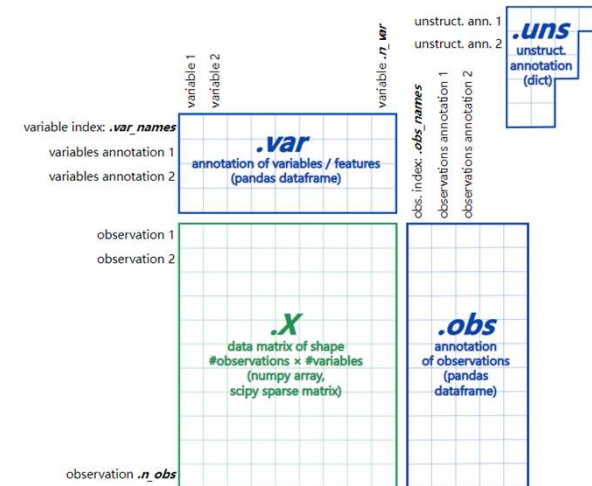
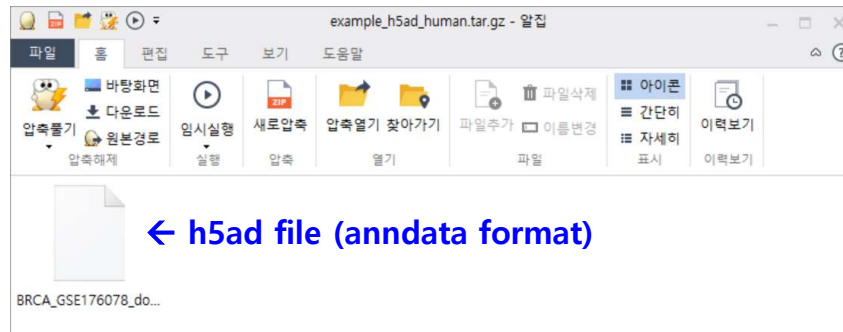
meta\_data

- 1열: 인덱스 열
- 2열: sample
- 3열: condition
- 4열~: optional items

- Count 데이터 파일과 메타 데이터 파일의 인덱스 열은 일치해야 함.
- Condition 열의 조건들을 대상으로 DEG, GSEA, cell-cell interaction 비교가 수행됨

# Input data formatting (3) h5ad format

## 1. 압축된 예제 데이터셋을 알집으로 열어 보면



## 2. h5ad file contents

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

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

AnnData object with n\_obs x n\_vars = 12000 x 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.