

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

윤석현

(주) 엠엘비아이랩 (MLBI Lab)

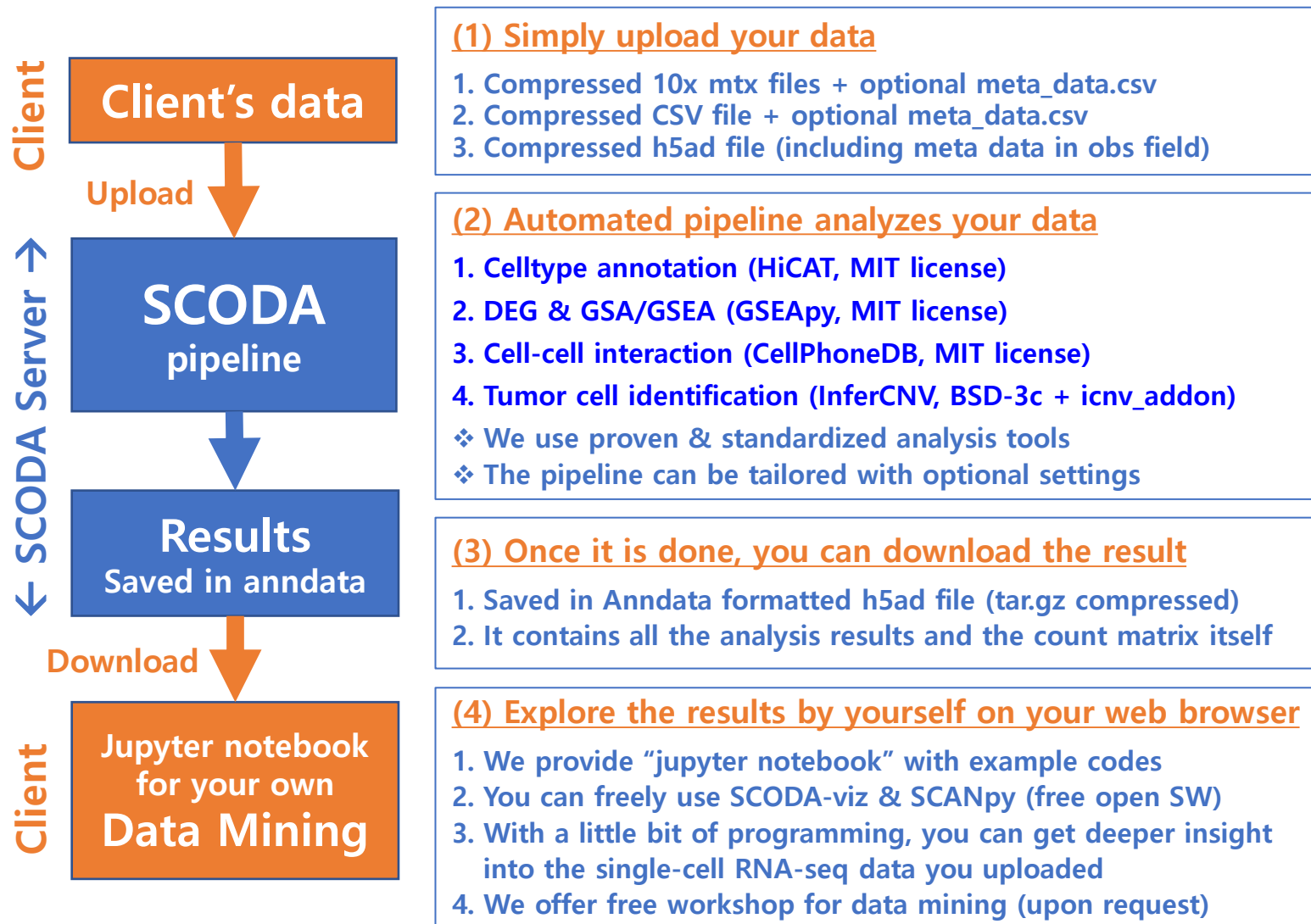
단국대학교 전자전기공학부

단국대학교 대학원 인공지능융합학과

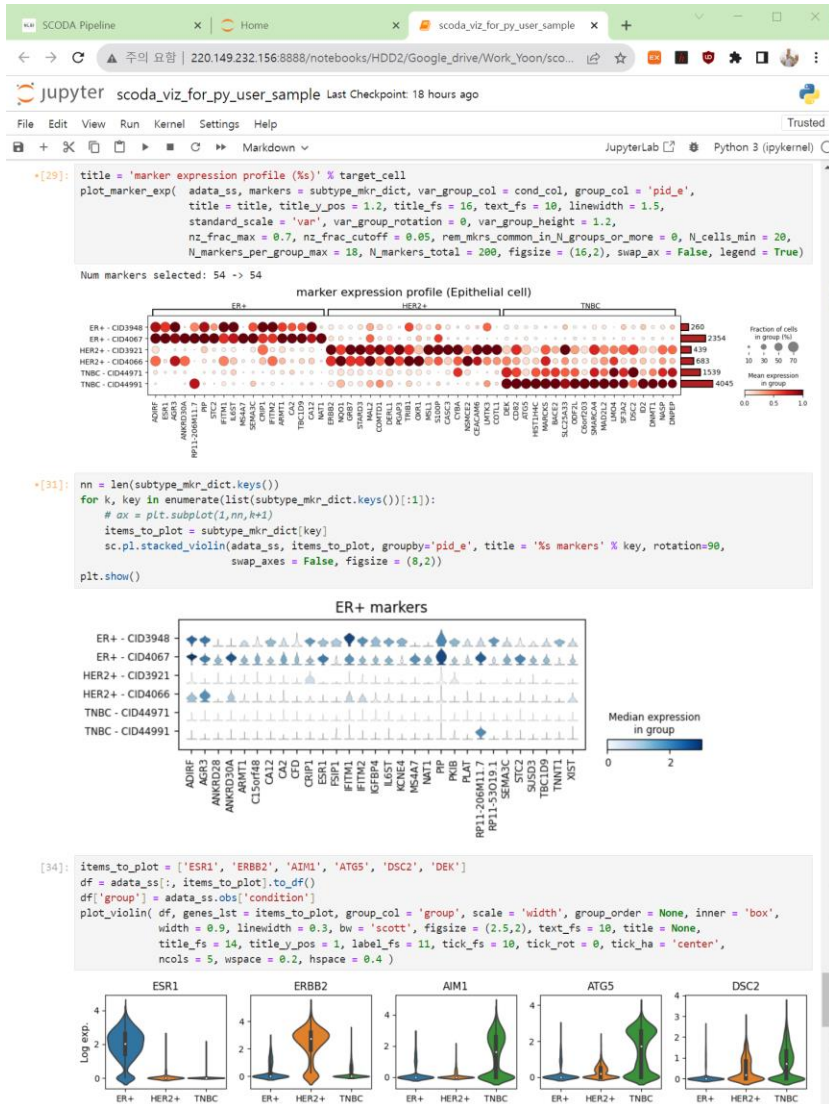
SCODA?

- ❑ Web-based, Fully-automated, all-in-one computing service for **S**ingle-**C**ell (transcript)**O**mics **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

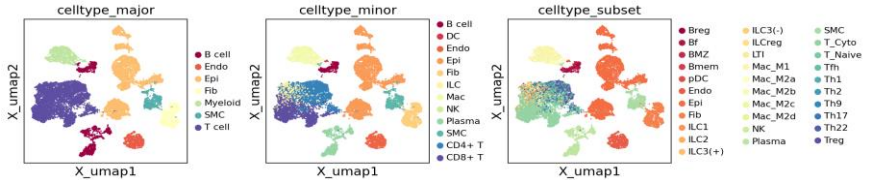
How SCODA works?



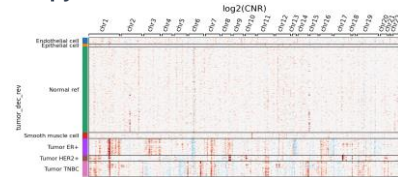
Visualization & data mining



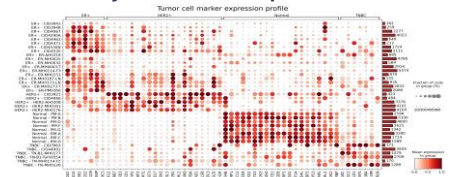
Cell type annotation in 3-level taxonomy



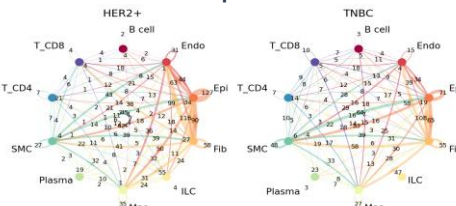
Copy Number Variation estimation



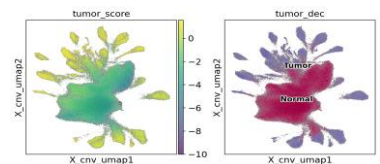
Discovery of condition specific markers



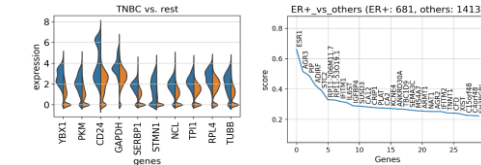
Cell-Cell interaction per condition



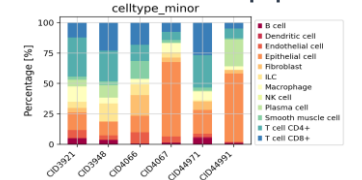
Tumor cell identification



DEG per cell-type, per condition

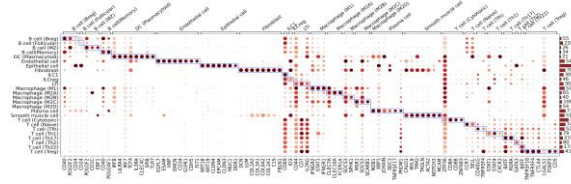


Cell population

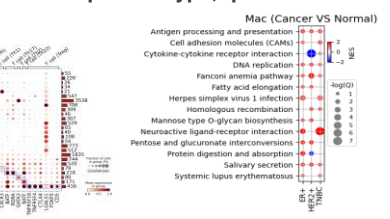


Verification of cell-type subset markers

subset marker expression profile



GSEA per cell type, per condition



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)

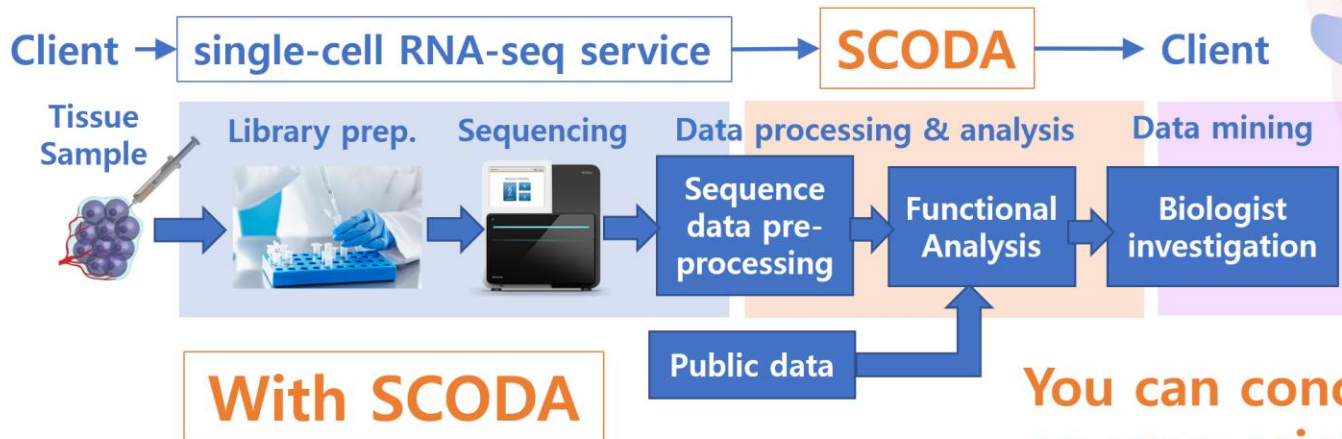
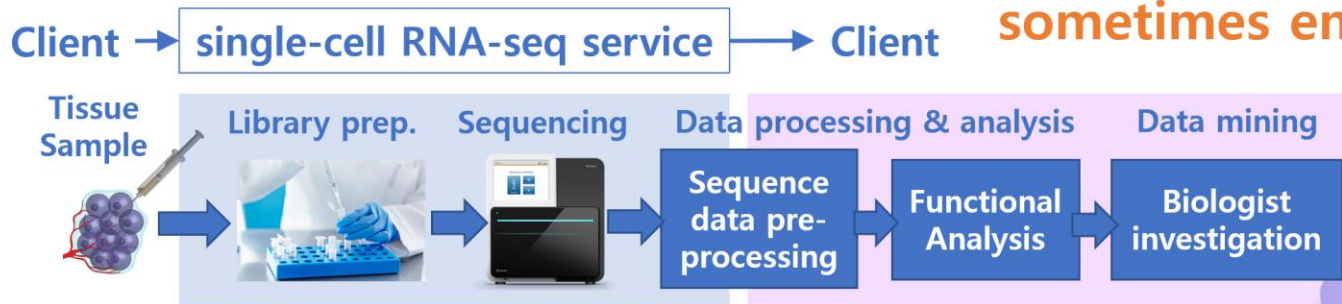
||

Single-cell RNA-seq
데이터 분석 전문가

Why SCODA?

Without SCODA

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

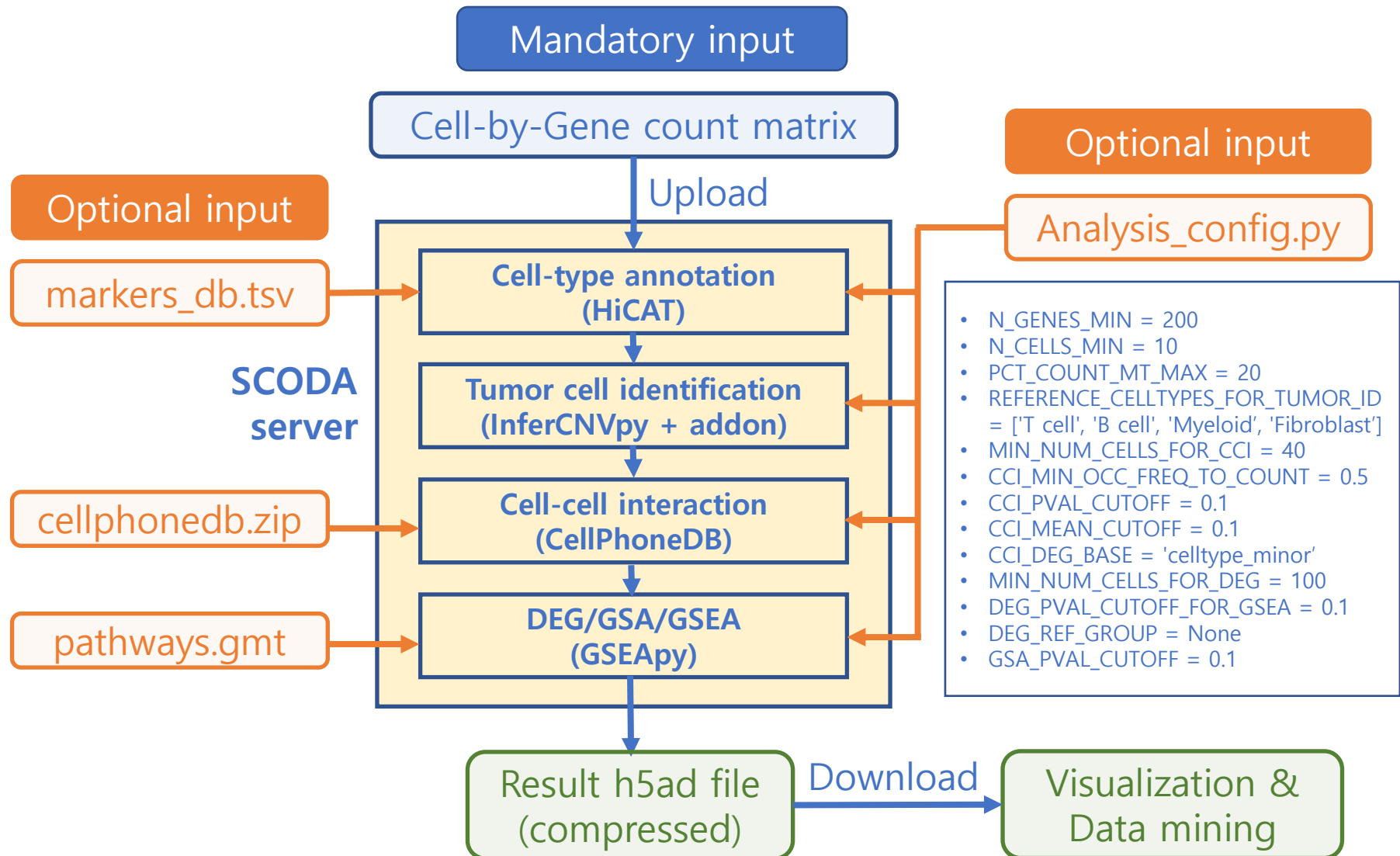
□ HiCAT

- J. Lee, M. Kim, K. Kang, CS Yang and S. Yoon, “Hierarchical cell-type identifier accurately distinguishes immune-cell subtypes enabling precise profiling of tissue microenvironment with single-cell RNA-sequencing,” Briefings in Bioinformatics, Jan. 2023. <https://doi.org/10.1093/bib/bbad006>

□ Studies using SCODA

- D. Hong, H. Kim, W. Yang, C. Yoon, M. Kim, CS Yang and S. Yoon, “Integrative analysis of single-cell RNA-seq and gut microbiome metabarcoding data elucidates macrophage dysfunction in mice with DSS-induced ulcerative colitis,” Communications Biology, June 2024. <https://doi.org/10.1038/s42003-024-06409-w>
- M. Kim, W. Yang, D. Hong, HS Won, S. Yoon, “A Retrospective View of the Triple-Negative Breast Cancer Microenvironment: Novel Markers, Interactions, and Mechanisms of Tumor-Associated Components Using Public Single-Cell RNA-Seq Datasets,” Cancers, Mar. 2024. <https://doi.org/10.3390/cancers16061173>
- JS Kim, HK Kim, M. Kim, S. Jang, E. Cho, S. Mun, J. Lee, D. Hong, S. Yoon and CS Yang,, “Colon-Targeted eNAMPT-Specific Peptide Systems for Treatment of DSS-Induced Acute and Chronic Colitis in Mouse,” Antioxidants, Nov. 2022. <https://doi.org/10.3390/antiox11122376>

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.

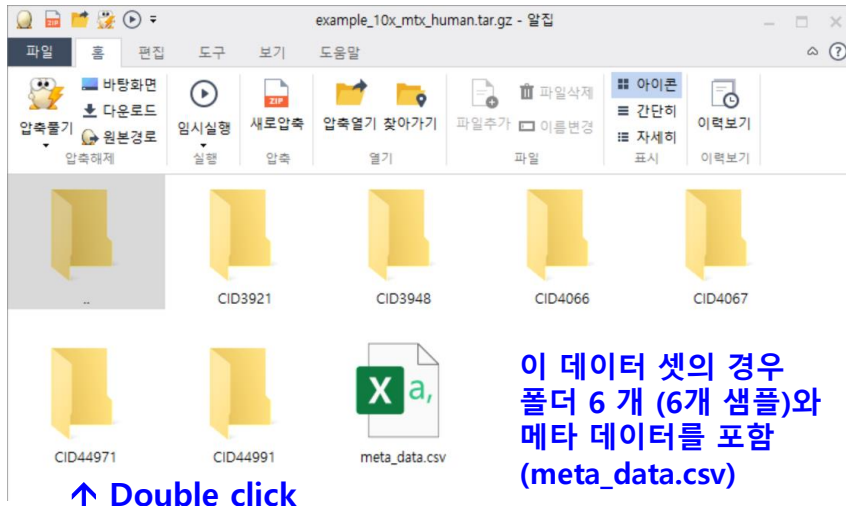
Input dataset formatting for using SCODA

Input data formatting (1) 10x_mtx

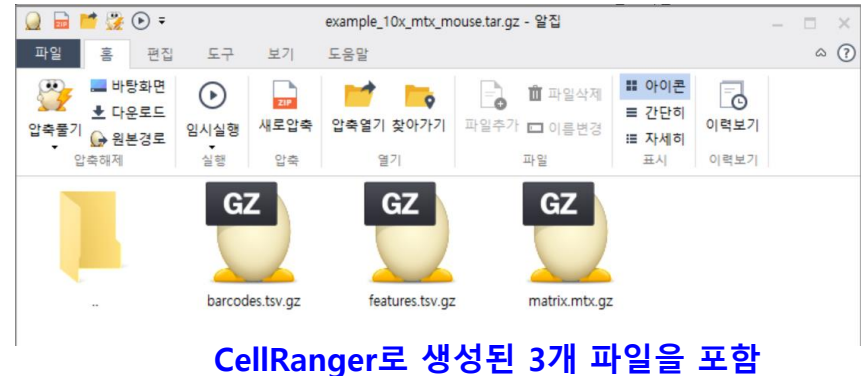
1. 압축된 예제 데이터셋을 알집으로 열어 보면



2. 메인 폴더를 더블 클릭해서 열면



3. 각 데이터 폴더를 더블 클릭해서 열면



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

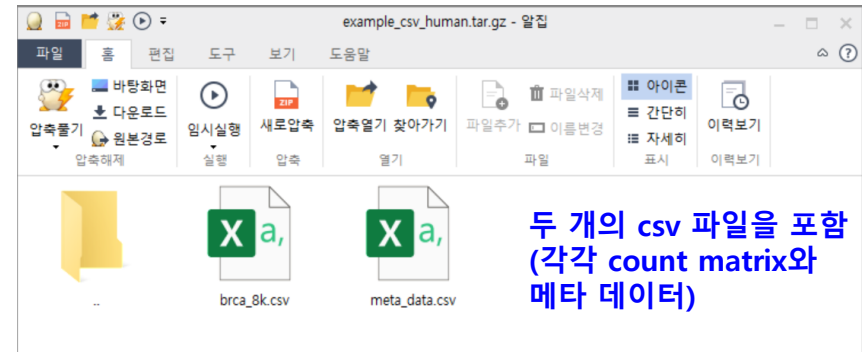
- 인덱스 열의 각 인덱스는 폴더명과 일치해야 함.
- Sample name과 인덱스가 동일할 필요는 없음.
- Condition 열의 조건들을 대상으로 DEG, GSEA, cell-cell interaction 비교가 수행됨

Input data formatting (2) csv format

1. 압축된 예제 데이터셋을 알집으로 열어 보면



2. 메인 폴더를 더블 클릭해서 열면



3. 데이터 csv 파일

Hugo symbol

Count matrix

Cell barcode (cell ID)

4. meta_data.csv 파일

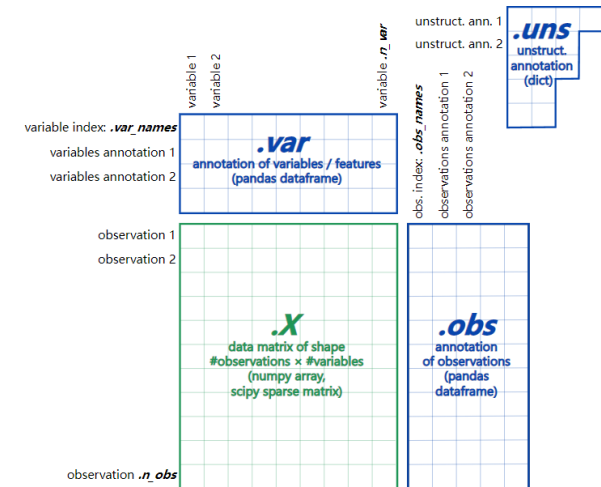
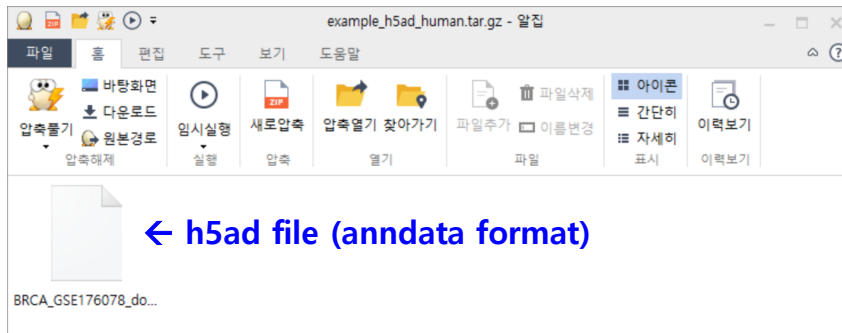
Cell barcode (cell ID)

- 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 × 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.