Background

Single-cell RNA-seq
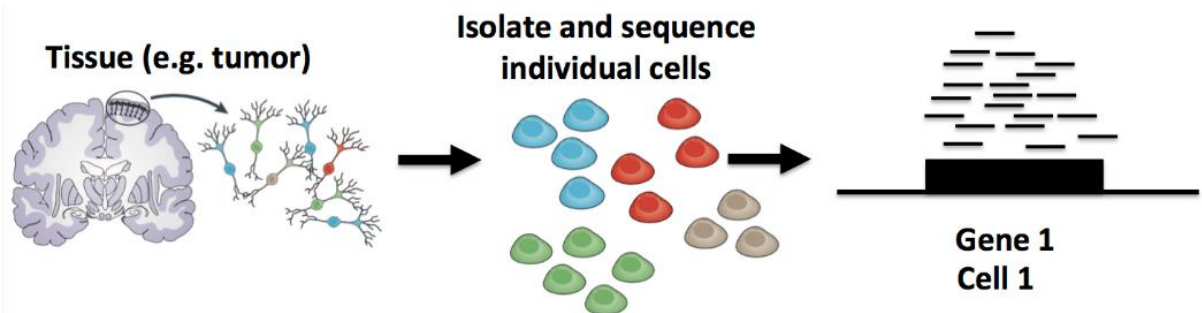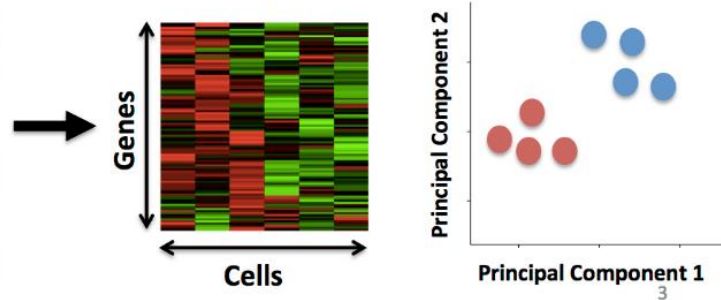
Single-cell proteomics

translation

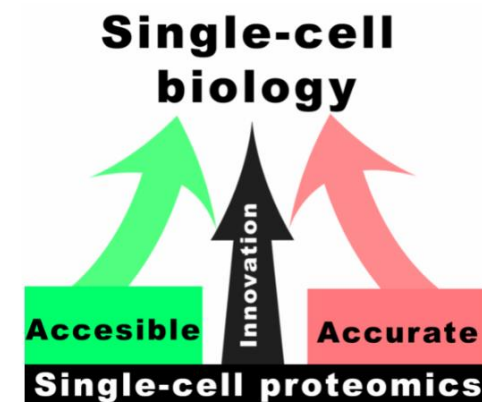# Existing single-cell proteomic processing pipeline
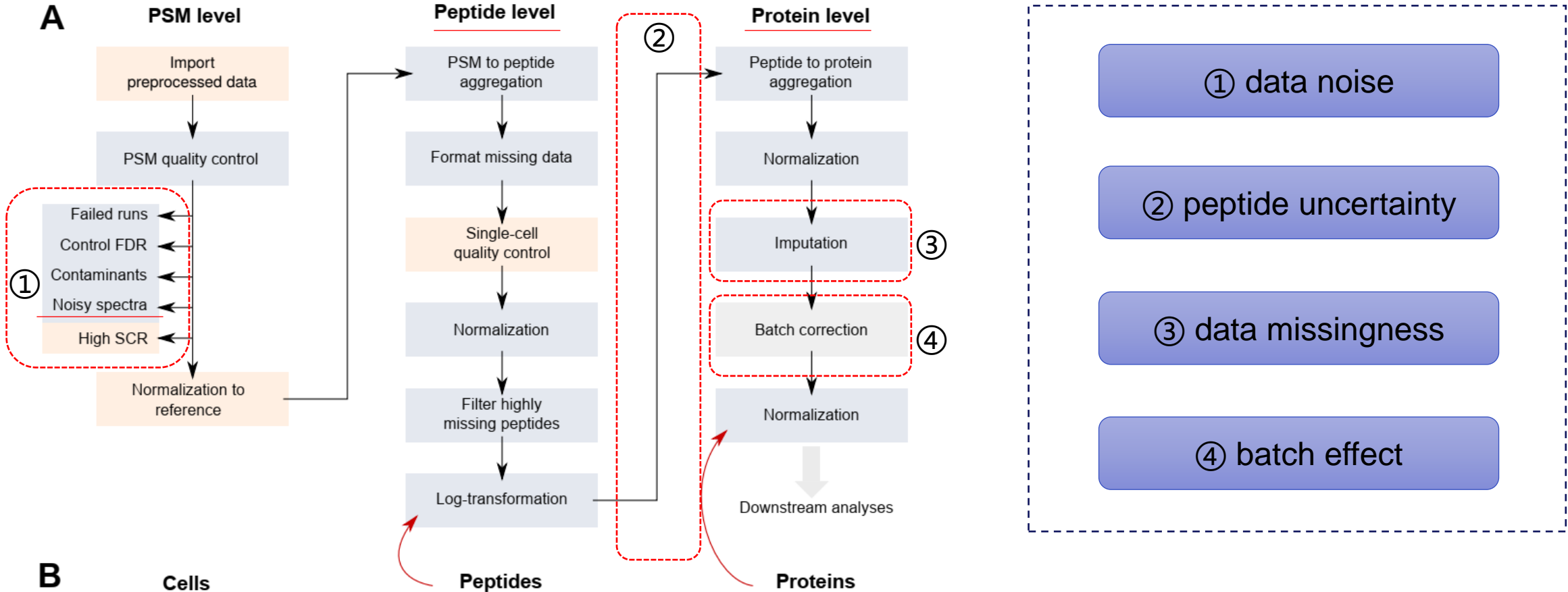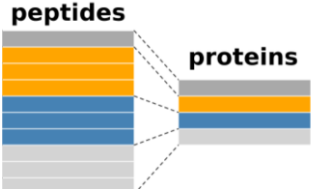
[1] Specht H, Emmott E, et al. Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. Genome biology, 2021
[2] https://uclouvain-cbio.github.io/scp/articles/scp.html

Method

➢ A versatile Deep Graph Contrastive Learning Framework (scPROTEIN) was developed for Single-cell Proteomics to tackle with this set of problems together

• stage 1: peptide uncertainty estimation
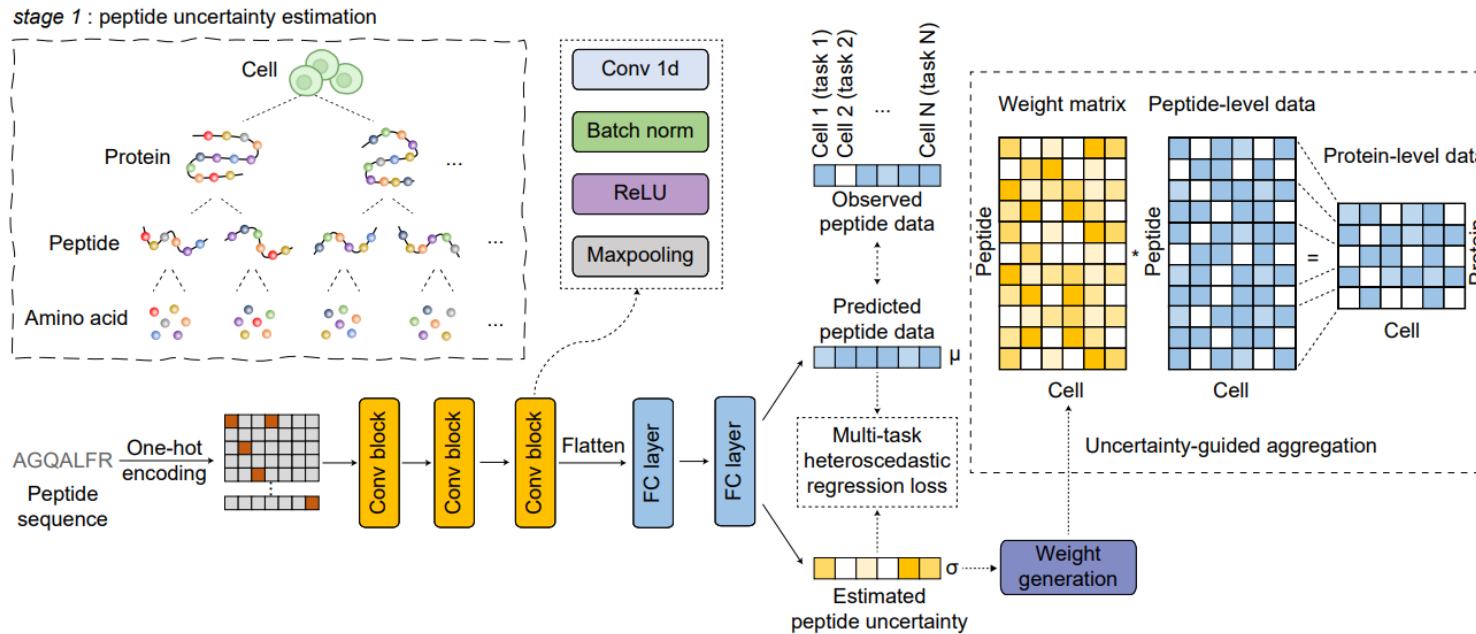
◆ Multi-task heteroscedastic regression model

• multi-task:

estimate different uncertainties of all cells

• heteroscedastic:

estimate different uncertainties across different peptides

◆ Uncertainty-guided peptide aggregation

Method

- stage 2: unsupervised cell embedding learning

stage 2 : unsupervised cell embedding learning
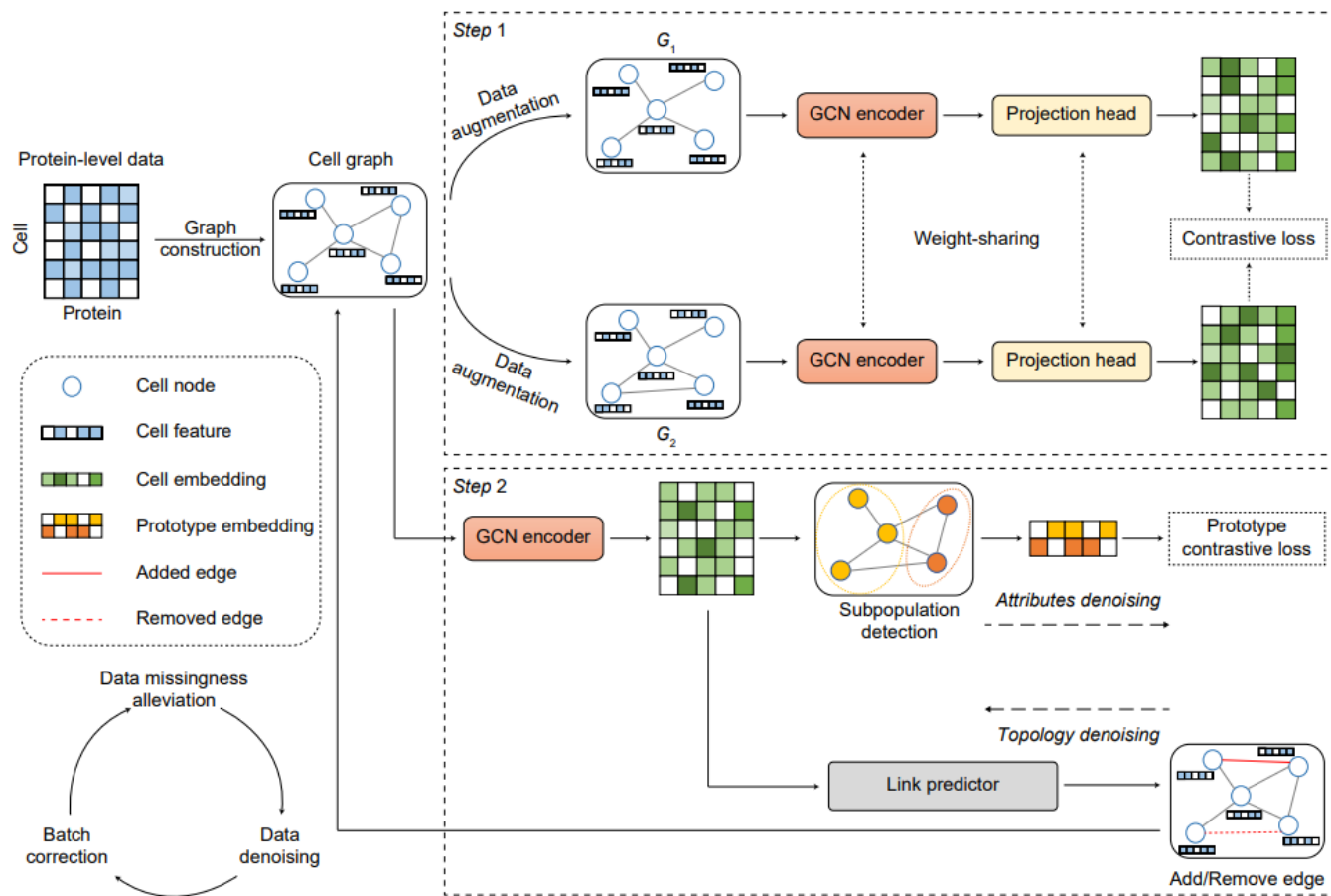
◆ Graph contrastive learning

- Message passing process among neighborhood can help alleviate the data missingness
- Batch effect can be implicitly alleviated by aligning the semantic information of the same cell type through contrastive loss

◆ Attribute-topology alternative denoising module

- Two denoising modules are alternated to mitigate the noise problem in the proteomic profile.

Method

stage 3: downstream analysis with learned cell embedding
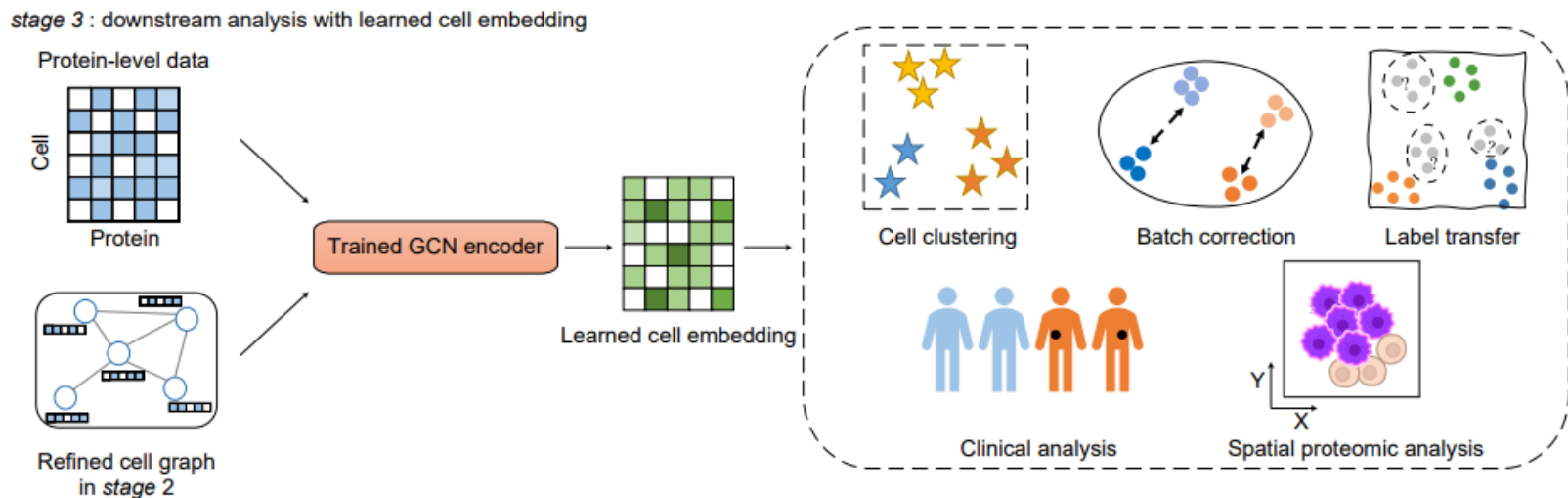
- stage 3: downstream analysis with learned embedding

The learned versatile embedding can be applied in a variety of downstream tasks:

- ✓ cell clustering
- ✓ batch correction
- ✓ label transfer
- ✓ clinical analysis
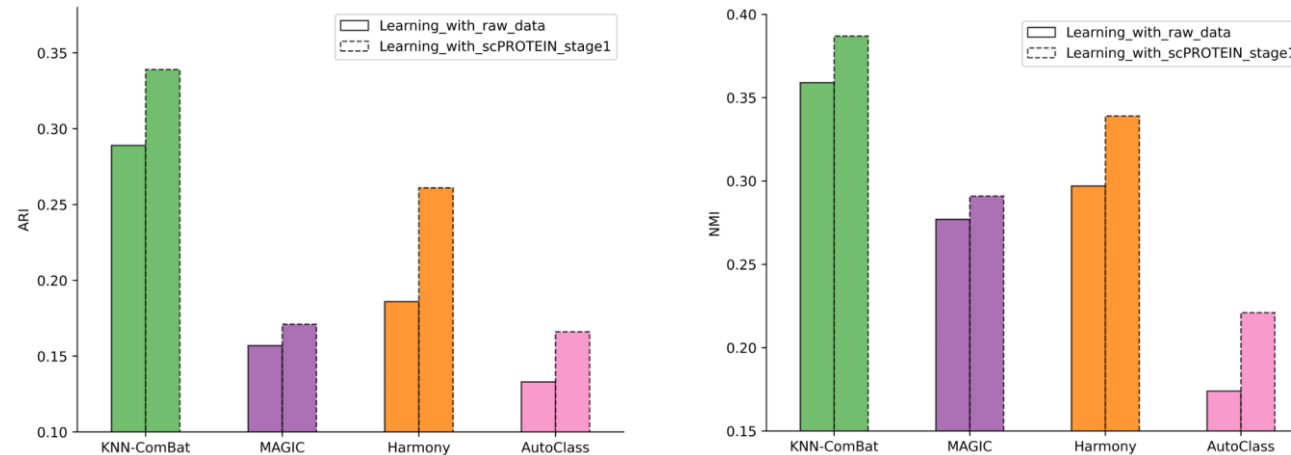- ✓ spatial proteome analysis

Results

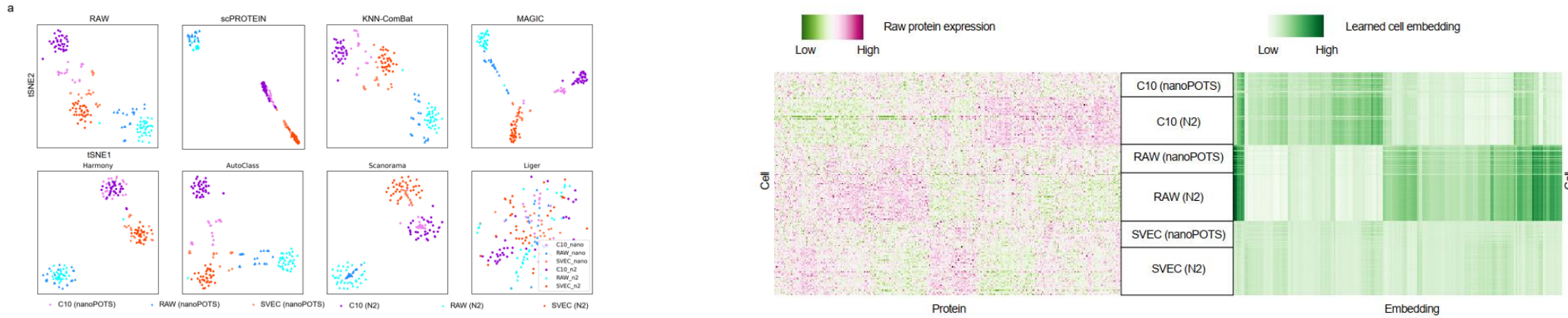➤ Cell clustering performance comparison on SCoPE2_Specht (1490 cells, 3042 proteins, two cell types)

ARI   NMI   ASW   PS
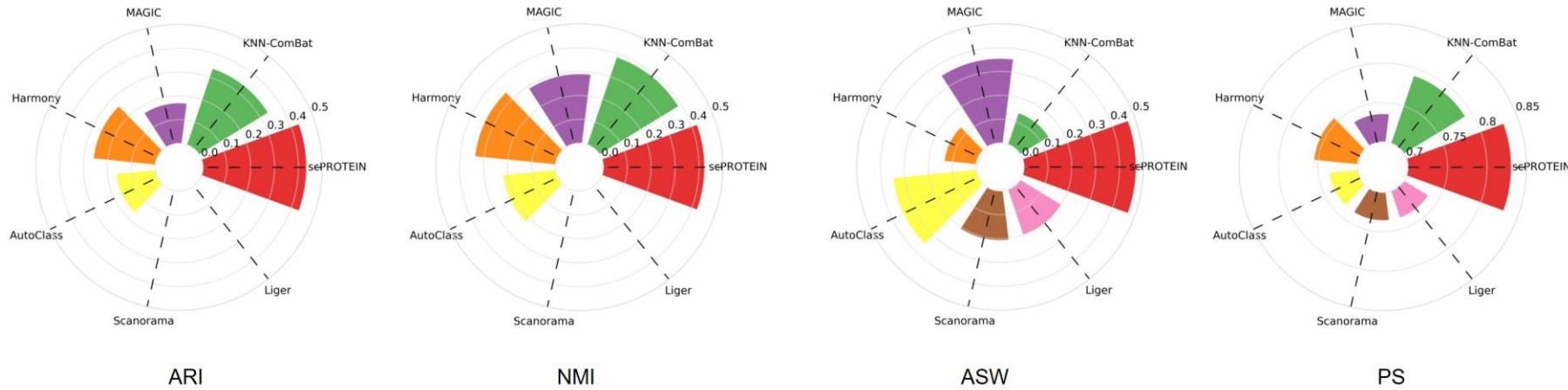
➤ Other methods with/wo scPROTEIN stage1

utilize the protein-level data from scPROTEIN stage1 as input for other methods

Results

➤ Integrate N2 dataset (108 cells and 1068 proteins) with nanoPOTS dataset (61 cells and 1225 proteins)

➤ Integrate SCoPE2_Leduc (163 cells and 1647 proteins) and plexDIA (164 cells and 1242 proteins) dataset

These two datasets have different cell types, task is more challenging.
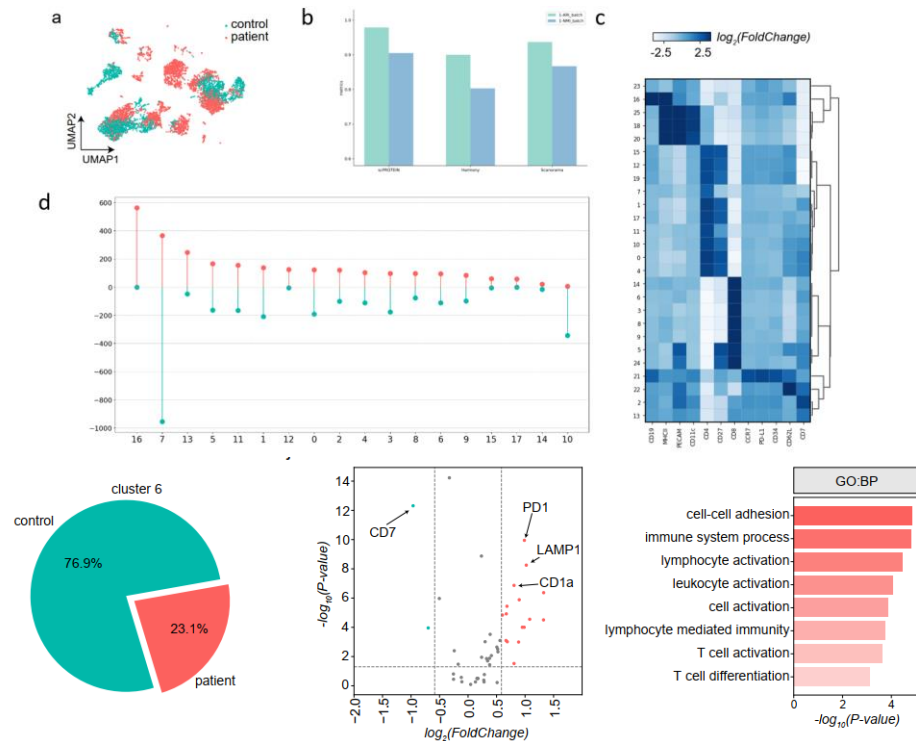
➤ Label transfer on N2 and nanoPOTS
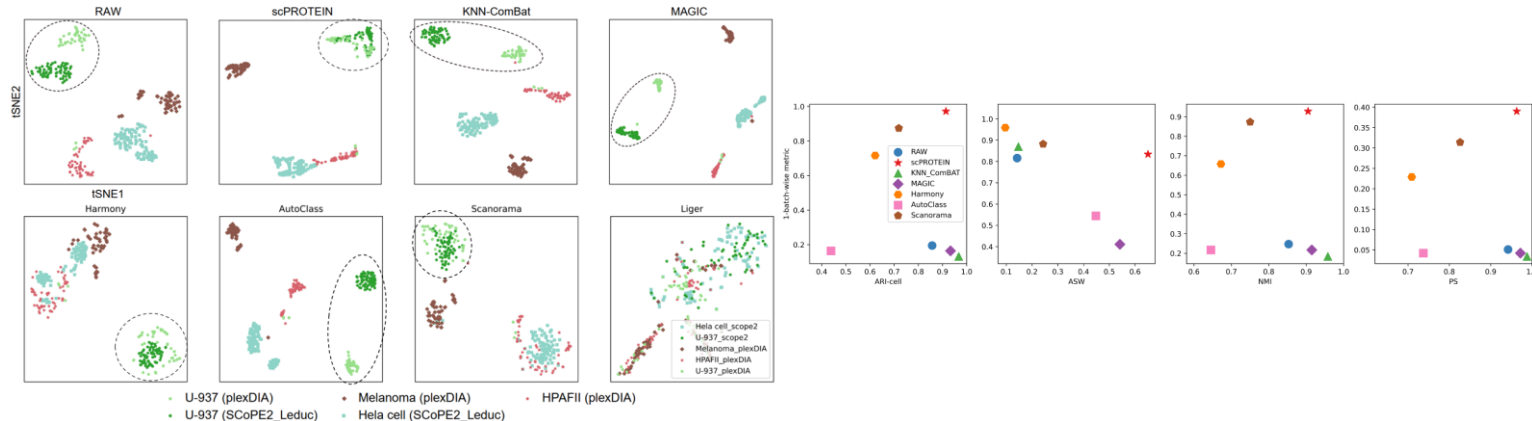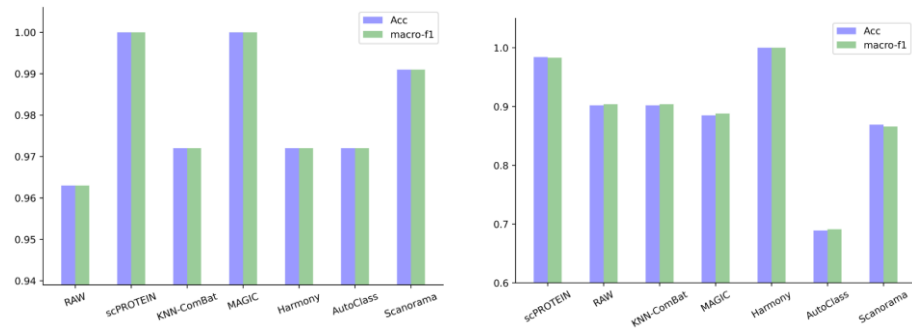
Results

- Clinic proteomics data analysis

- ECCITE-seq dataset (6500 cells from a healthy donor and 6500 cells from a patient with CTCL)

- Correct batch effect between two donors

- Find PD1, which is vital marker of CTCL

- Spatial proteomics data analysis

- BaseITMA dataset includes 281 patients with breast cancer, with 38 marker proteins.

- Use spatial location for graph construction

- The cell embedding can be used to analysis tumor/non-tumor slices