



Swiss Institute of
Bioinformatics

SINGLE-CELL TRANSCRIPTOMICS WITH R

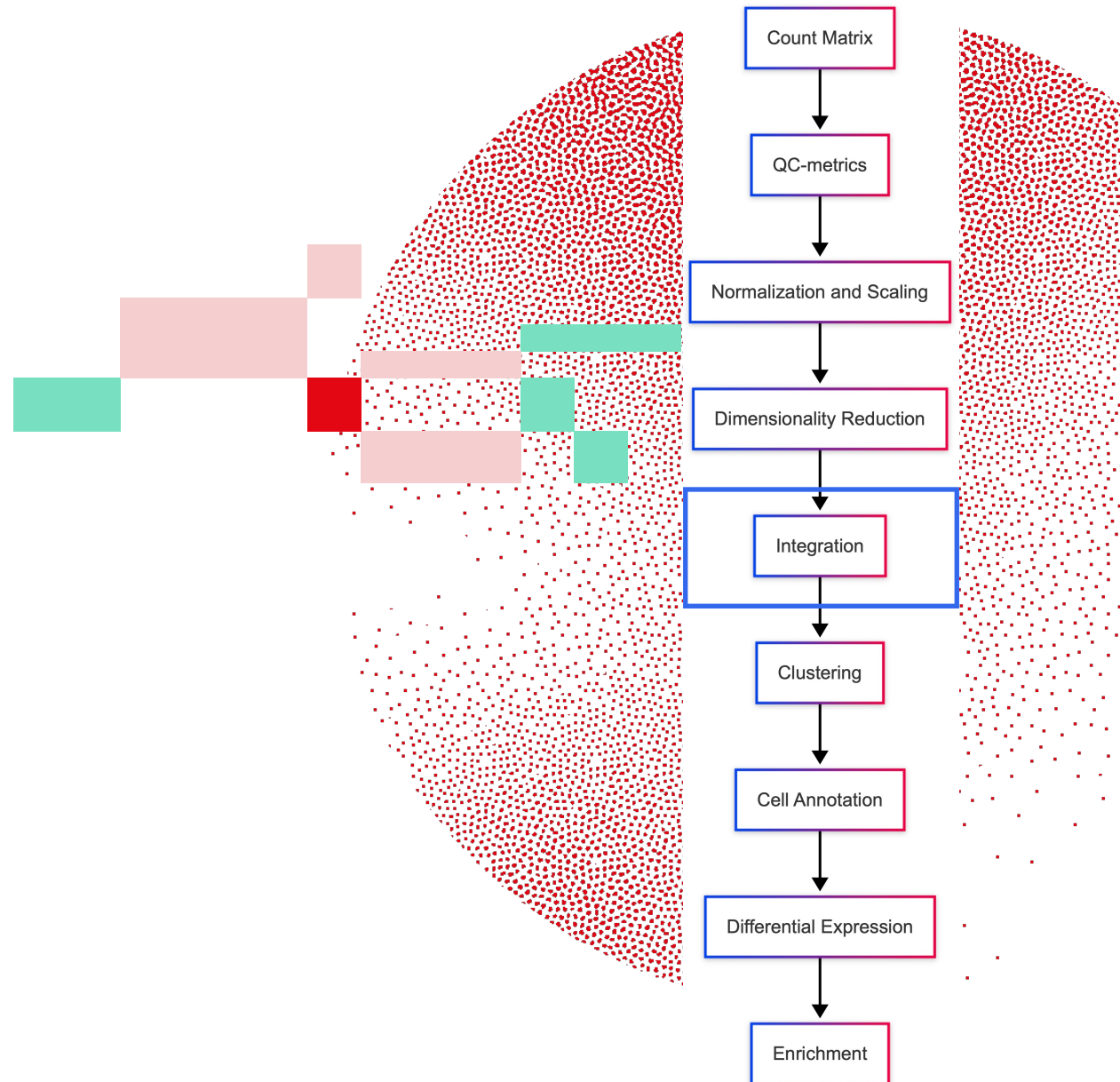
Integration analysis

Deepak Tanwar

March 18-20, 2025

Adapted from previous year courses

Feedback from Geert van Geest



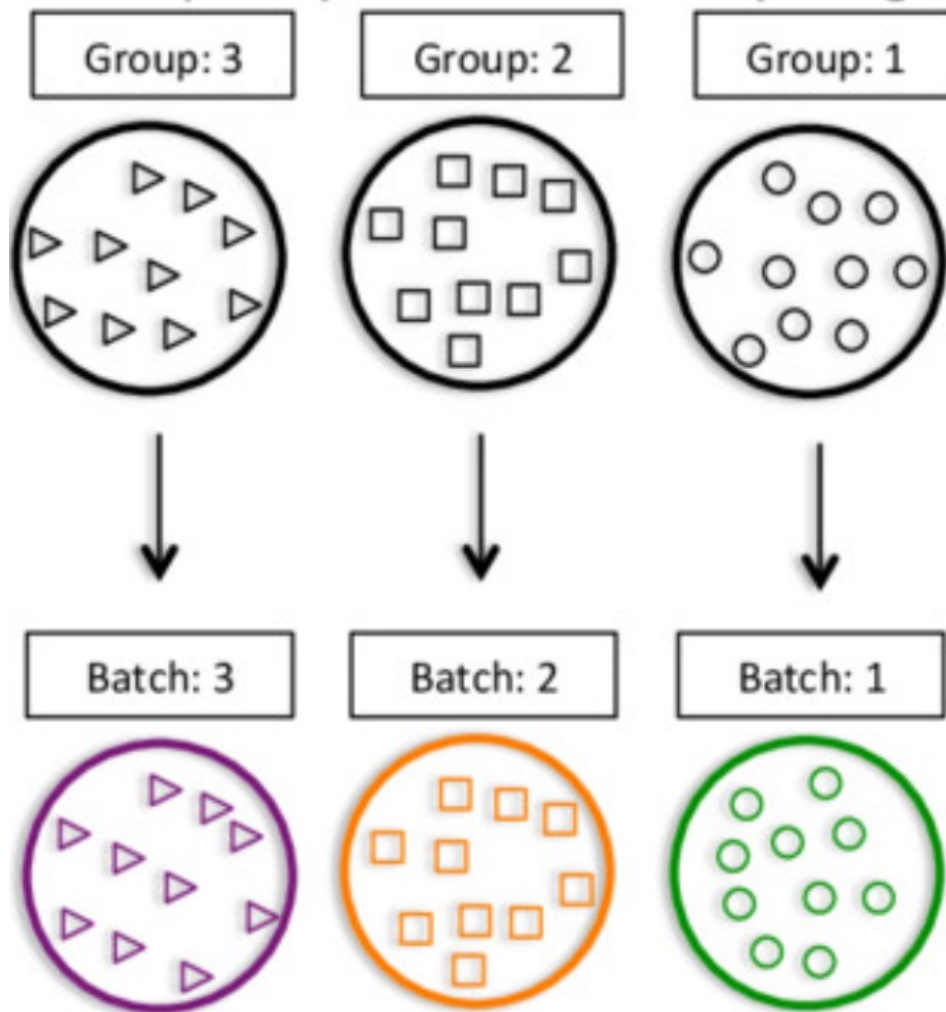
Learning objectives

Understand the importance of experimental design

Identify scenarios where integration is necessary for data analysis

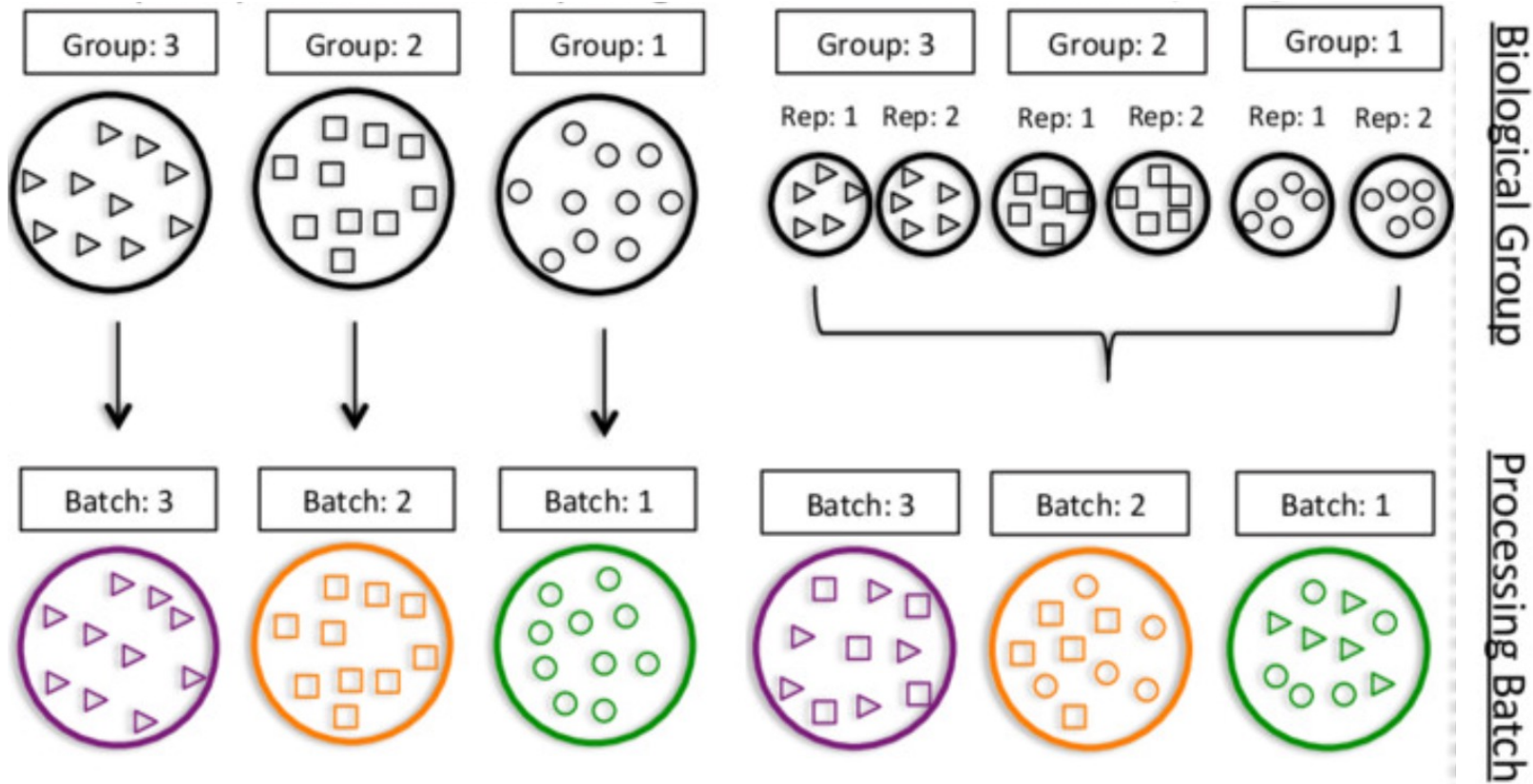
Apply canonical correlation analysis (CCA) for integrating datasets

Experimental design matters

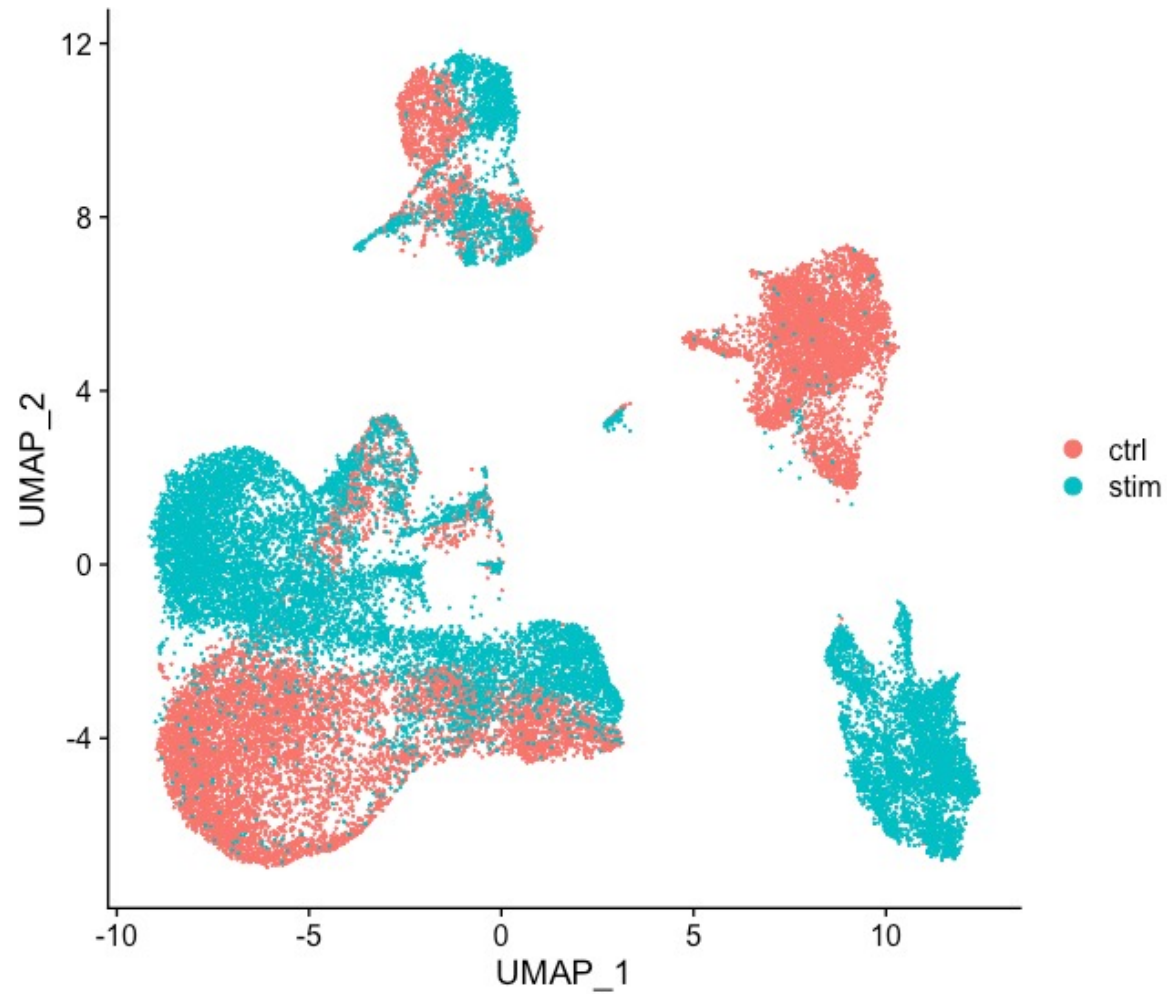


What changes would you make here to make the experimental design more optimal?

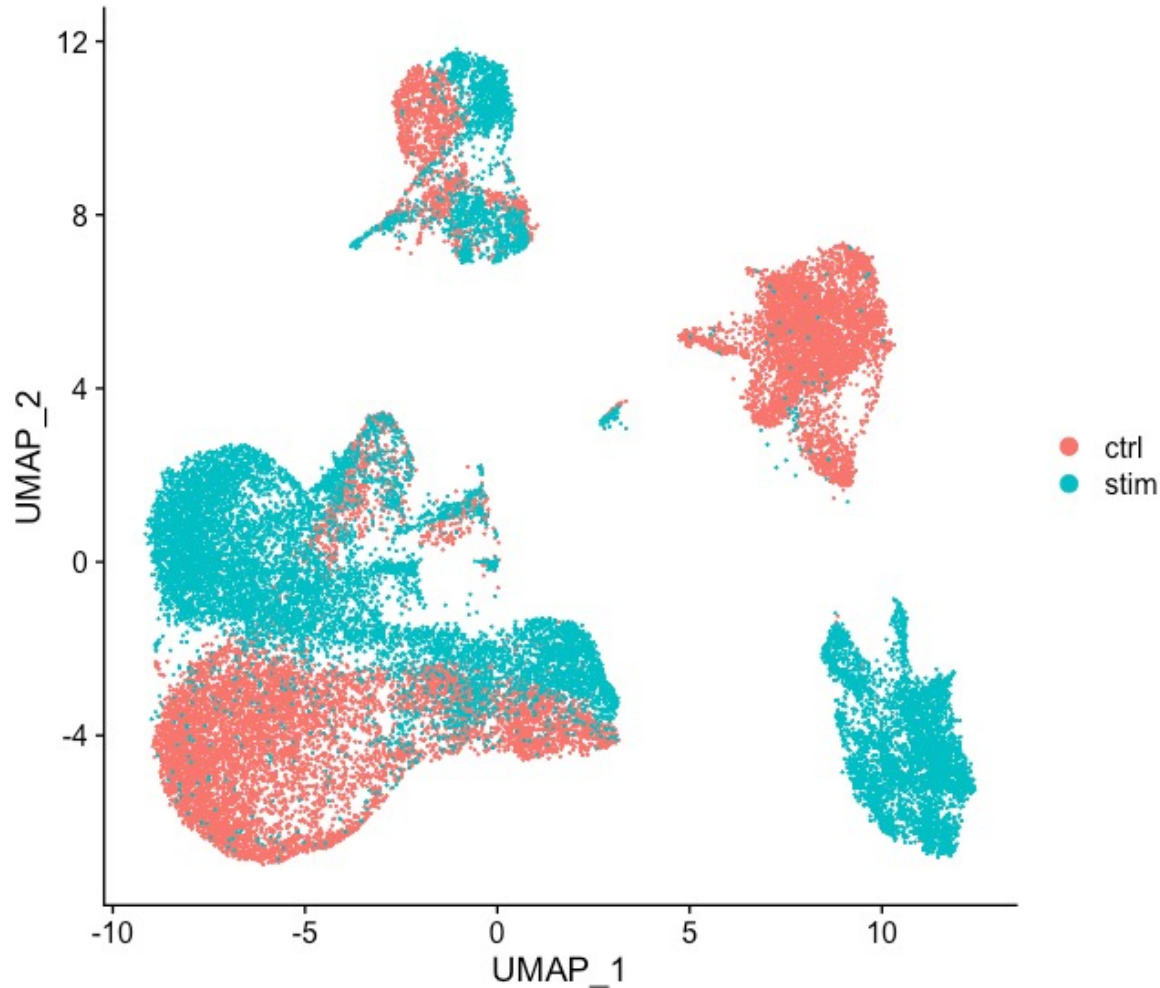
Experimental design matters



Exercise: Identify problem in this plot

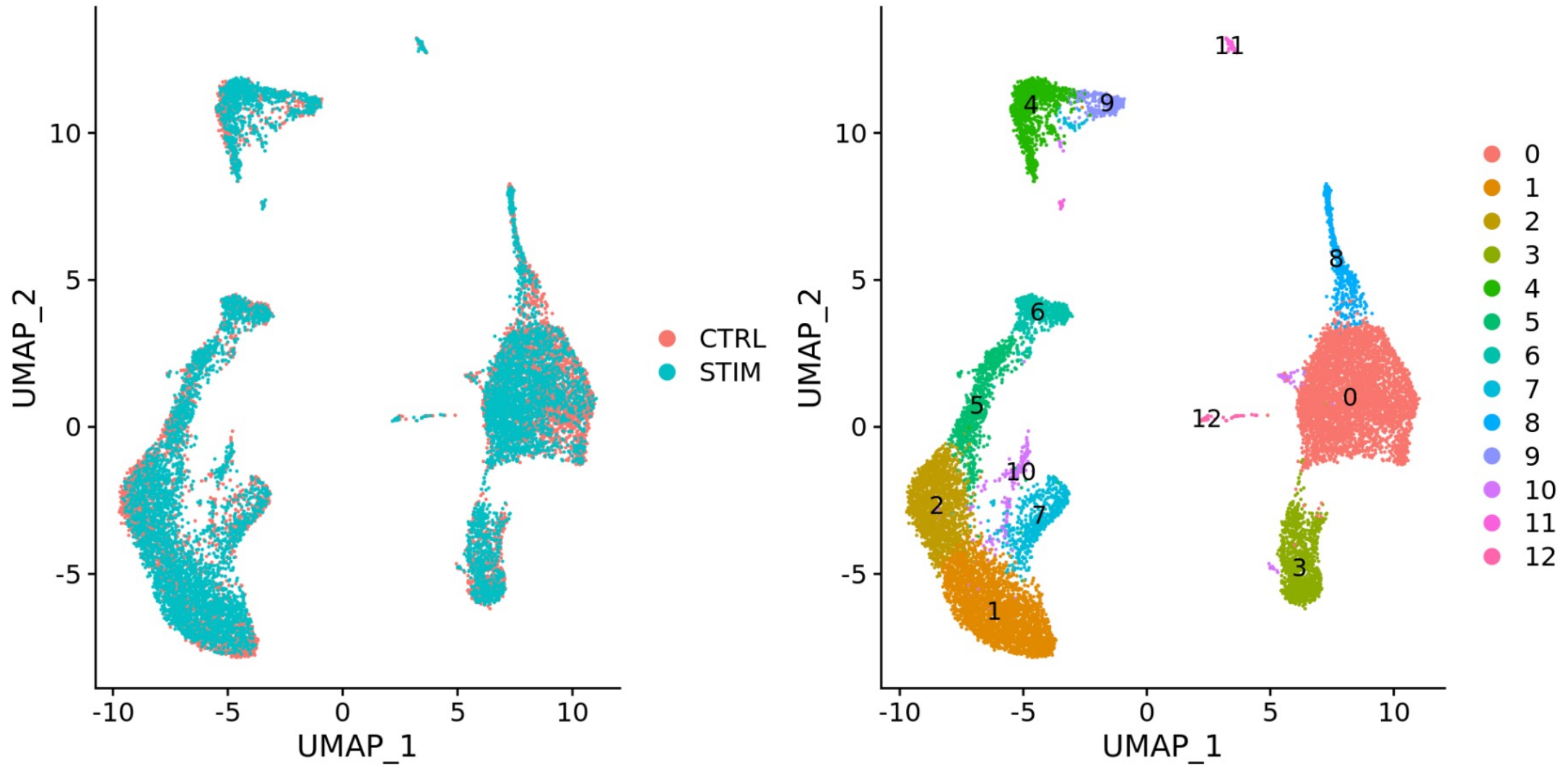


Exercise: Identify problem in this plot



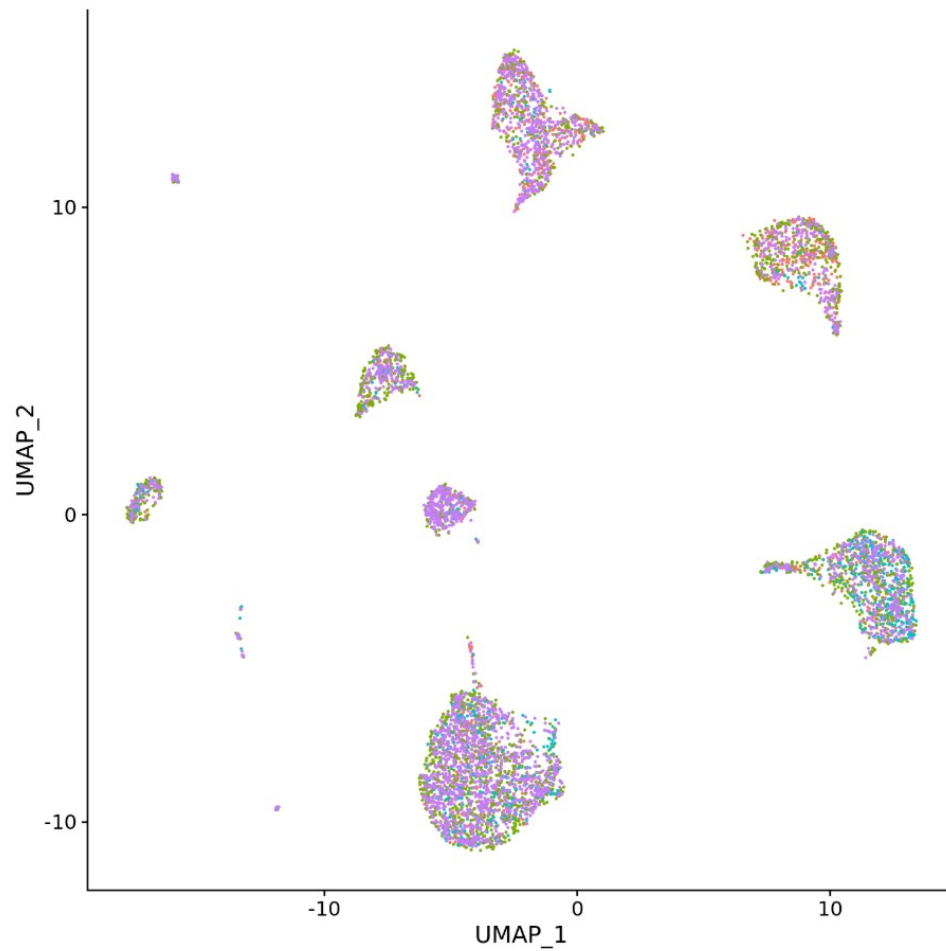
- **Explore the data:** do not just always perform integration because you think there might be differences
- If cells cluster by **sample, condition, batch, dataset, modality**, performing integration can help align cells across the groups to greatly improve the clustering and the downstream analyses.

Example scenarios for integration: **conditions**

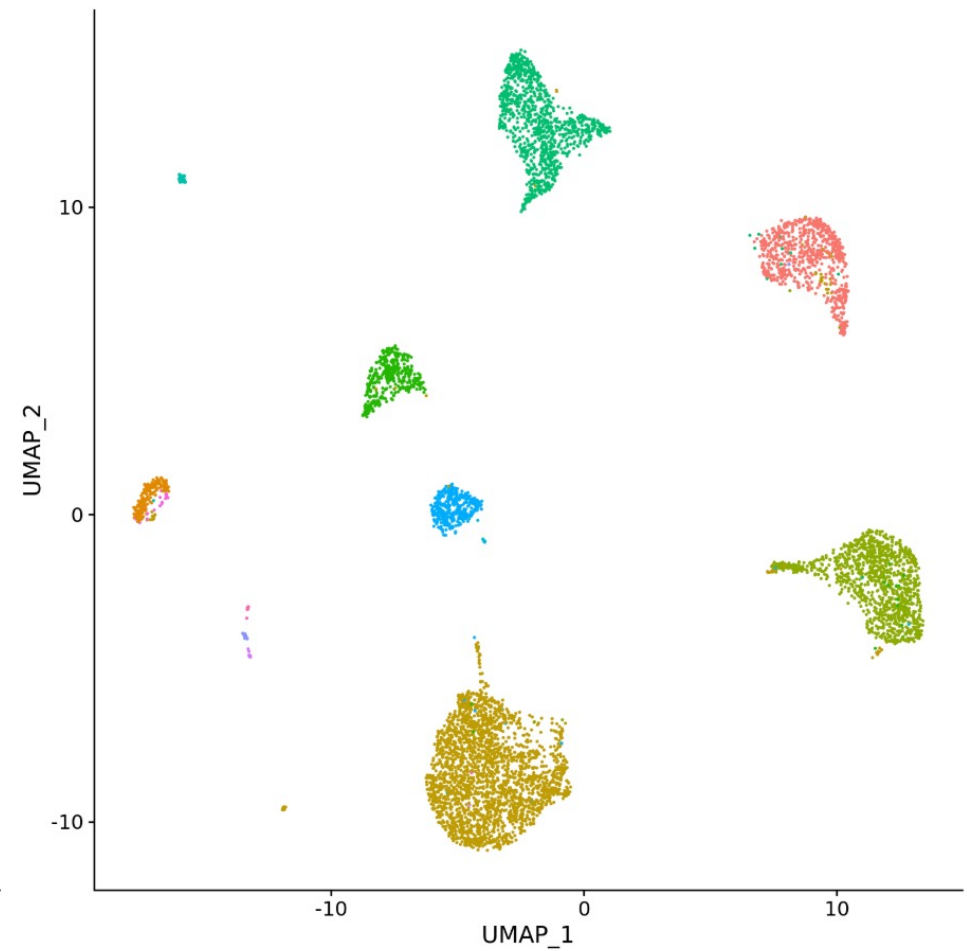


Example scenarios for integration: **datasets**

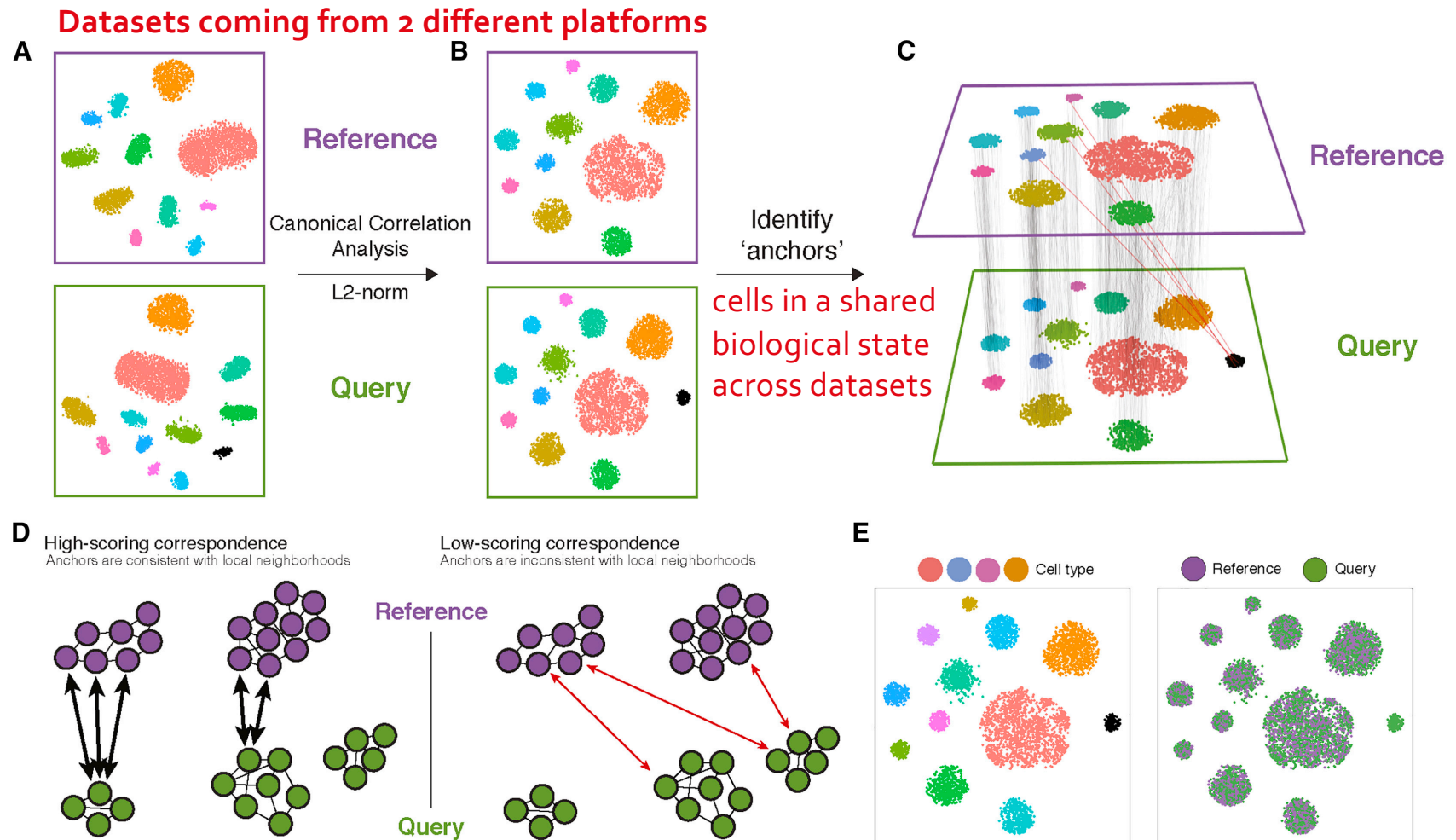
● celseq ● celseq2
● fluidigm1 ● smartseq2



● acinar ● activated_stellate ● alpha ● beta ● delta
● ductal ● endothelial ● epsilon ● gamma ● macrophage
● mast ● quiescent_stellate ● schwann



Integration using CCA: canonical correlation analysis



scores to compute "correction" vectors for each query cell,
transforming its expression so it can be jointly analyzed

Quiz

In which condition will you perform integration?

- A) When cells cluster by sample
- B) When cells cluster by condition
- C) When cells cluster by batch
- D) When cells cluster by dataset
- E) None of the above
- F) All of the above

Summary

Experimental Design Matters: Optimize design to improve data quality and analysis

Integration Scenarios:

- Conditions: Compare different experimental conditions
- Datasets: Analyze data from different platforms together

Integration Using CCA:

- Align cells across groups to improve clustering and downstream analyses
- Compute correction vectors for each query cell to transform its expression for joint analysis



sib.swiss

