Haghverdi et al., "Batch effects in single-cell RNA sequencing data are corrected by matching mutual nearest neighbours", Nat. Biotechnol., 36(5):421-427, 2018

Session Intro

The session looks at RNA-sequencing in the context of single-cell gene expression analysis, aka scRNA-seq. In particular, we discuss the paper by Haghverdi et al., which describes a key idea that underlies many approaches to batch-effect correction of scRNA-seq datasets. The highly cited paper has a rather intuitive and seemingly attractive idea: Mutual nearest neighbours across two datasets should correspond to the same cell types and therefore are excellent anchor points for defining corrections or mappings from one batch to the other.

Session Plan

I am dividing the session into three parts as given below. However, I leave each presenting team to decide on what they want to talk about (i.e., it is perfectly ok to leave out some topics/points/details and/or include other topics/points/details.) Also, the presenting team need not just make presentations; they are encouraged to figure out how to engender more class interactions and lead discussions.

Part I, What scRNA-Seq is:

This part deals with background knowledge of scRNA-Seq. These wikipedia pages are ok starting points, https://en.wikipedia.org/wiki/Single-cell_transcriptomics and https://en.wikipedia.org/wiki/Single_cell_sequencing.

Some suggestions on things to present:

- Overview of how scRNA-Seq works and its applications

- Details of some scRNA-Seq technologies such as Chromium 10x

- Common issues encountered in scRNA-seq datasets

Presentation team #5: KHOOI XIN ZHE, QIN HANGYU

Total time limit: 15 minutes (presentation) + 5 minutes (audience questions.) Total slide count: 10 slides max.

Part II, the paper by Haghverdi et al.

This part presents the Haghverdi et al. paper itself. We want to know the key technical details and the key messages.

Presentation team #6: DIBYADIP CHATTERJEE, GAO TIANYU, LIU NIAN

Total time limit: 15 minutes (presentation) + 5 minutes (audience questions.) Total slide count: 10 slides max.

Part III, Possible points for discussion

This part discusses the Haghverdi et al. paper, hopefully in depth. We want to know whether there is any methodological issue, any doubt on the conclusions/key messages, any suggestion for improving the paper. Some pointers for discussion include:

- The mutual nearest neighbour method is defined on a pair of data batches. Can you apply it when there are more than two batches?

- The mutual nearest neighbour method assumes mutual nearest neighbours correspond to the same cell type. What will happen if there are some cell types which are unique to a batch?

Presentation team #2: MALAIKA AFRA TAJ, DAI YUHE, LEE JIANYI DAVID

Total time limit: 15 minutes (presentation) + 5 minutes (audience questions.) Total slide count: 10 slides max.