



Solution to practice questions

Xiaochen Zhang

Jiadong Mao

MIG

Exercise 1

Two PCs are not enough to see any interesting structures. This is because 2 PCs are not enough to capture interesting variations in data. As we use more and more PCs, we see more detailed structures within the data.

However, if we use too many (say 100) then the structures we see might be spurious (and the downstream computation will also be slower). In this toy dataset, it seems that the choices of 20, 30, 40, 50 PCs gave similar UMAP visualisation.

Exercise 4

The annotation provided by the Azimuth is better as Azimuth uses a larger reference data set.

This means that Azimuth can learn the characteristics of each cell type to obtain accurate cell type annotations.

Exercise 6

Data points for the same cell type from different sequencing platforms changed from two clusters to one cluster.

The algorithm successfully discovered and removed batch effects.