1. What is the difference between the two PBMC datasets?

PBMC-Zheng and PBMC-Kang come from the same cell types but not the same tissue.

1. How do you process your datasets?

I got the data from Zheng and Kang, and I followed the processing process of scGen to get a side-by-side fair comparison. We first filtered the cell with a minimum of 500 expressed genes. Then we removed the cell type with unstable labels (the megakaryocyte cells). After that, we log-transformed the data to make the training process more smoothly. Finally, we paired the control group and the perturbed group divided by each cell type for the style transferring process.

1. What is ?

R^2 is the R^2 value between the model's prediction and the ground truth. The closer the R^2 is to 1, the closer the prediction is to the ground truth, so the better the model makes the prediction.

1. What method did you use to divide the training and testing dataset?

I randomly split the dataset into training sets and testing sets. The ratio between these two sets is 8:2.

1. Did you do an ablation study?
2. Where do you get your data?
3. How to filter your data?
4. Why do you choose to filter your data in such a way? Will it filter out important data?
5. Could your model predict the mutation of genes?
6. How do you choose the coefficient of the loss function?

The coefficient of the loss function means to balance the scale of each loss. For example, the KL loss is much larger than Generated loss or style loss, then the coefficient of KL loss is much lower than that of style loss. The ratio between the coefficient of KL to Generated loss is about 1: 1000.