Project proposal: Modelling of scRNA data

Kasper Rasmussen

June 2020

1 Description

This project will focus on scRNA data from the perspective of a computer science student with only basic understanding of the biological domain, thus focusing mainly on mathematical and algorithmic aspects. It will centre on the part of a pipeline which would involve denoising, imputation and dimensionality reduction fx for use in identifying cell types.

- Describing scRNA-seq and the processing leading to the vector representations of cell transcriptomes. Describing the biological and technical reasons why the data is noisy and in need of imputation.

- Surveying some of the methods applied to scRNA data for denoising, imputation, or for learning lower-dimensional manifolds. This includes learning about the mathematical theory involved.

- Implementing one or more of the methods and considering alternative variants.

- Describing common downstream tasks and performing some of these tasks. This could be clustering.

- Discussing measures of performance and how to benchmark the methods. Performing these measurements and weighing the pros and cons of different methods.

References

[1] Azizi, Elham, et al. "Bayesian inference for single-cell clustering and imputing." *Genomics and Computational Biology* 3.1 (2017): e46-e46.

[2] Christopher Heje Grønbech, Maximillian Fornitz Vording, Pascal Timshel, Casper Kaae Sønderby, Tune Hannes Pers, Ole Winther, scVAE: Variational auto-encoders for single-cell gene expression data, *Bioinformatics*, btaa293, https://doi.org/10.1093/bioinformatics/btaa293

[3] Eraslan, Gökcen, et al. "Single-cell RNA-seq denoising using a deep count autoencoder." *Nature communications* 10.1 (2019): 1-14.

[4] Gong, Wuming, et al. "DrImpute: imputing dropout events in single cell RNA sequencing data." *BMC bioinformatics* 19.1 (2018): 220.

[5] Kiselev, V.Y., Andrews, T.S. & Hemberg, M. "Challenges in unsupervised clustering of single-cell RNA-seq data." *Nat Rev Genet* 20, 273–282 (2019). https://doi.org/10.1038/s41576-018-0088-9

[6] Lin, P., Troup, M. & Ho, J.W. CIDR: "Ultrafast and accurate clustering through imputation for single-cell RNA-seq data." *Genome Biol* 18, 59 (2017). https://doi.org/10.1186/s13059-017-1188-0

[7] Lopez, Romain, et al. "Deep generative modeling for single-cell transcriptomics." *Nature methods* 15.12 (2018): 1053-1058.

[8] Talwar, Divyanshu, et al. "AutoImpute: Autoencoder based imputation of single-cell RNA-seq data." *Scientific reports* 8.1 (2018): 1-11.

[9] Van Dijk, David, et al. "Recovering gene interactions from single-cell data using data diffusion." *Cell* 174.3 (2018): 716-729.