Research Week 2021

Creating interpretable and informative low-dimensional representations of cell-state from sequencing data using visible neural network variational autoencoders (VNNVAE)

Nathaniel Evans[1]*, Shannon McWeeney[2,†]

[1]evansna@ohsu.edu, [2]mcweeney@ohsu.edu

*Division of Bioinformatics and Computational Biology, Department of Medical Informatics & Clinical Epidemiology, Oregon Health & Science University, Portland, Oregon, United States of America

†Knight Cancer Institute, Oregon Health & Science University, Portland, Oregon, United States of America, Division of Bioinformatics and Computational Biology, Department of Medical Informatics & Clinical Epidemiology, Oregon Health & Science University, Portland, Oregon, United States of America, Oregon Clinical and Translational Research Institute, Oregon Health & Science University, Portland, Oregon, United States of America

Keywords

Deep learning, dimensionality reduction, precision oncology, next-generation sequencing

Abstract

Next generation sequencing (NGS) and single-cell sequencing technologies have enabled unparalleled resolution of the cell's molecular machinery; however, gleaning accurate knowledge from sequencing data is often stymied by high-dimensionality, measurement noise and biological complexity. Numerous methods have been proposed to address this by mapping high-dimensional inputs to informative low-dimensional representations. One such dimensionality reduction method is the variational autoencoder (VAE), which attempts to learn the probability distribution of given data through a low-dimensional latent variable, and has been shown to competitively separate cell types and to characterize functional differences of cell state[1,2]. A limitation of deep learning methods like VAE's are 1) understanding the internal operations of the model (interpretability) and 2) the generalization of a model to new data (overfitting). Recent work by the Ideker lab has proposed Visible Neural Networks[3,4] (VNN) to address these issues by constraining feature interactions using literature curated biological hierarchies. In this talk, we present our preliminary results implementing a novel VAE model that incorporates prior knowledge through a VNN to create low-dimensional cell-state representations using bulk and single-cell RNA expression features. We hypothesize that this will reduce model complexity (number of parameters) while maintaining or improving model performance and creating a biologically relevant low-dimensional representation of sequencing data. Successful execution of this research will provide an interpretable and informative deep learning dimensionality reduction algorithm. This work was originally motivated by challenges in precision oncology, but may have application in many biological domains.

References

1. Lin E, Mukherjee S, Kannan S. A deep adversarial variational autoencoder model for dimensionality reduction in single-cell RNA sequencing analysis. BMC Bioinformatics. 2020 Feb 21;21(1):64.

2. Grønbech CH, Vording MF, Timshel P, Sønderby CK, Pers TH, Winther O. scVAE: Variational auto-encoders for single-cell gene expression data. bioRxiv. 2019 May 21;318295.

3. Ma J, Yu MK, Fong S, Ono K, Sage E, Demchak B, et al. Using deep learning to model the hierarchical structure and function of a cell. Nat Methods. 2018 Apr;15(4):290–8.

4. Kuenzi BM, Park J, Fong SH, Sanchez KS, Lee J, Kreisberg JF, et al. Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells. Cancer Cell [Internet]. 2020 Oct 22 [cited 2020 Oct 26]; Available from: http://www.sciencedirect.com/science/article/pii/S1535610820304888z