Relative Entropy Used to Determine the Divergence of Cells in Single Cell RNA Sequence Data Analysis

Authors : An Chengrui, Yin Zi, Wu Bingbing, Ma Yuanzhu, Jin Kaixiu, Chen Xiao, Ouyang Hongwei

Abstract : Single cell RNA sequence (scRNA-seq) is one of the effective tools to study transcriptomics of biological processes. Recently, similarity measurement of cells is Euclidian distance or its derivatives. However, the process of scRNA-seq is a multi-variate Bernoulli event model, thus we hypothesize that it would be more efficient when the divergence between cells is valued with relative entropy than Euclidian distance. In this study, we compared the performances of Euclidian distance, Spearman correlation distance and Relative Entropy using scRNA-seq data of the early, medial and late stage of limb development generated in our lab. Relative Entropy is better than other methods according to cluster potential test. Furthermore, we developed KL-SNE, an algorithm modifying t-SNE whose definition of divergence between cells Euclidian distance to Kullback–Leibler divergence. Results showed that KL-SNE was more effective to dissect cell heterogeneity than t-SNE, indicating the better performance of relative entropy than Euclidian distance. Specifically, the chondrocyte expressing Comp was clustered together with KL-SNE but not with t-SNE. Surprisingly, cells in early stage were surrounded by cells in medial stage in the processing of KL-SNE while medial cells neighbored to late stage with the process of t-SNE. This results parallel to Heatmap which showed cells in medial stage were more heterogenic than cells in other stages. In addition, we also found that results of KL-SNE tend to follow Gaussian distribution compared with those of the t-SNE, which could also be verified with the analysis of scRNA-seq data from another study on human embryo development. Therefore, it is also an effective way to convert non-Gaussian distribution to Gaussian distribution and facilitate the subsequent statistic possesses. Thus, relative entropy is potentially a better way to determine the divergence of cells in scRNA-seq data analysis.

Keywords : Single cell RNA sequence, Similarity measurement, Relative Entropy, KL-SNE, t-SNE
Conference Title : ICBI 2017 : International Conference on Bioinformatics
Conference Location : Bangkok, Thailand
Conference Dates : October 26-27, 2017