World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:14, No:01, 2020

Improving Cell Type Identification of Single Cell Data by Iterative Graph-Based Noise Filtering

Authors : Annika Stechemesser, Rachel Pounds, Emma Lucas, Chris Dawson, Julia Lipecki, Pavle Vrljicak, Jan Brosens, Sean Kehoe, Jason Yap, Lawrence Young, Sascha Ott

Abstract : Advances in technology make it now possible to retrieve the genetic information of thousands of single cancerous cells. One of the key challenges in single cell analysis of cancerous tissue is to determine the number of different cell types and their characteristic genes within the sample to better understand the tumors and their reaction to different treatments. For this analysis to be possible, it is crucial to filter out background noise as it can severely blur the downstream analysis and give misleading results. In-depth analysis of the state-of-the-art filtering methods for single cell data showed that they do, in some cases, not separate noisy and normal cells sufficiently. We introduced an algorithm that filters and clusters single cell data simultaneously without relying on certain genes or thresholds chosen by eye. It detects communities in a Shared Nearest Neighbor similarity network, which captures the similarities and dissimilarities of the cells by optimizing the modularity and then identifies and removes vertices with a weak clustering belonging. This strategy is based on the fact that noisy data instances are very likely to be similar to true cell types but do not match any of these wells. Once the clustering is complete, we apply a set of evaluation metrics on the cluster level and accept or reject clusters based on the outcome. The performance of our algorithm was tested on three datasets and led to convincing results. We were able to replicate the results on a Peripheral Blood Mononuclear Cells dataset. Furthermore, we applied the algorithm to two samples of ovarian cancer from the same patient before and after chemotherapy. Comparing the standard approach to our algorithm, we found a hidden cell type in the ovarian postchemotherapy data with interesting marker genes that are potentially relevant for medical research.

Keywords : cancer research, graph theory, machine learning, single cell analysis
Conference Title : ICABSB 2020 : International Conference on Applied Biomedical Statistics and Biomathematics
Conference Location : Paris, France
Conference Dates : January 23-24, 2020

Open Science Index, Biomedical and Biological Engineering Vol:14, No:01, 2020 publications.waset.org/abstracts/114686.pdf