BITS :: Call for Abstracts 2021 - Oral communication

Type	Oral communication
Session	Gene regulation, transcriptomics and epigenomics
Title	Stardust: spatial tranScripTomics data analysis through space awARe modularity optimization baseD clUSTering.
All Authors	Giovanni Motterle(1), Luca Alessandrì(2), Eva Viesi(1), Vincenzo Bonnici(1), Francesca Cordero(3), Marco Beccuti(3), Raffaele Calogero(2) and Rosalba Giugno(1)

Affiliation

(1) Department of Computer Science, University of Verona
(2) Department of Molecular Biotechnology and Health Sciences, University of Torino
(3) Department of Computer Science, University of Torino

Motivation

Spatial transcriptomics (ST) is able to combine stained tissue images with high-throughput spatially resolved RNA sequencing (Ståhl et al., 2016, Asp et al., 2020). A very important part of single cell transcriptomics is clustering, which represents the key element to identify the cell sub-population structure, characterizing a healthy or a disease tissue. ST data provides an extra layer of knowledge, with respect to conventional single cell RNAseq platforms, since ST keeps track of physical localization of the cells in the tissue under analysis. Notably, conventional software packages for single cell transcriptomics analysis perform clustering taking into account only the transcriptional similarity and do not handle the spatial information provided by ST data. Here we present Stardust, a clustering algorithm extension that combines transcriptional and positional information.

Methods

Seurat package (Butler et al., 2018) is one of the most used software for scRNAseq data analysis. It implements a network based algorithm for data clustering called Louvain algorithm (Blondel et al., 2008). Louvain algorithm encodes each element of a dataset as a node in a graph and then connects each node couple according to a pairwise measure of similarity based on the euclidean distance between transcriptional profiles. It applies a community detection step to the graph in order to generate the dataset partition. Stardust is implemented on top of the Seurat clustering algorithm, replacing the distance matrix computation. The new matrix is the result of the summation of two matrices. The first matrix is obtained from the pairwise euclidean distance between transcriptional profiles in PCA space (Jolliffe et al., 2016). The second matrix represents the pairwise euclidean distance between spatial positions of spots. Users can configure a parameter to control how much the space based measure weights on the overall measure. The proposed methodology can be incorporated in any existing single cell clustering methods. We used the rCASC R package (Alessandrì et al., 2019) to quantitatively measure the performance of ST clustering. The basic concept of the rCASC notion is that a good clustering should remain stable if a perturbation is applied to the dataset. A stability score is assigned to each spot depending on how many times that spot remained clustered with the same other spots in n different dataset perturbations. A statistical evaluation of the results is provided to assess if space contributes in a statistically significant way to the increase of the overall stability.

Results

We tested Stardust on 5 publicly available ST tissues provided by 10X Genomics (Human Lymph Node, Human Heart, Human Breast Cancer 1, Human Breast Cancer 2 and Mouse Kidney). In each dataset we computed 5 configurations of Stardust, obtained varying the weight given to space in the similarity matrix computation, from no space (which results correspond to plain Seurat) up to equal weight of space and transcriptional similarity. We observe that, in all datasets, space increases the stability scores distribution. Comparing the results of the best space-transcript configuration of Stardust with the ones of Seurat we discover that the increase in stability of Stardust is achieved mainly by merging clusters that are physically close in the tissue, by splitting a single cluster composed by different areas far away from each others or by rearranging the cluster identities assignment is a particular area.

Info

https://docs.google.com/document/d/1BOzj35frCcmrswztvDuEU4kF8024JjmpOo3XTEUlVBw/edit?usp=sharing

References:

Asp, M. et al. BioEssays 42.10 (2020)
Alessandrì, L. et al. GigaScience 8.9 (2019)
Blondel, Vincent D et al. Journal of Statistical Mechanics: Theory and Experiment 2008.10 (2008)
Butler, A. et al. Nature Biotechnology 36.5 (2018)
Jolliffe I.T. et al. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374.2065 (2016)
Ståhl et al. Science 353.6294 (2016)

Figure	
-	
Availability	https://docs.google.com/document/d/1yDE4J-XhZKBOtKYzHrEdeKHyL5gQx50yUzvyRaRaC0k/edit?usp=sharing

Corresponding Author

Name, Surname	Giovanni, Motterle
Email	giovanni.motterle@univr.it
Submitted on	30.04.2021