BITS :: Call for Abstracts 2022 - Oral communication

Type	Oral communication
Session	Spatial transcriptomics
Title	A new space-based method for downstream analysis of spatial transcriptomics data
All Authors	Simone Avesani*(1), Eva Viesi*(1), Luca Alessandrì(2), Giovanni Motterle(1), Vincenzo Bonnici(3), Marco Beccuti(4), Raffaele Calogero#(2), Rosalba Giugno#(1) * equal contributor # equal contributor

Affiliation

(1) Department of Computer Science, University of Verona, Verona
(2) Department of Molecular Biotechnology and Health Sciences, University of Turin, Turin
(3) Department of Mathematical, Physical and Computer Sciences, University of Parma, Parma
(4) Department of Computer Science, University of Turin, Turin

Motivation

Spatial transcriptomics (ST), which achieves the combination of whole transcriptome sequencing with spatial context information on a tissue section, is mainly performed by the 10x Genomics Visium [1-2], Slide-seqV2 [3] and Seq-Scope [4] technologies. Since these technologies keep track of the physical position of the captured cells or spots on a tissue, they provide additional knowledge to the one obtained with the standard single-cell workflow. A challenging task when analyzing ST data is clustering, which represents a powerful way to explore cell structures and substructures and characterize tissue pathology based on some distance metrics. By investigating the transcriptomic and spatial information of the datasets analyzed, we noticed that the relation between gene expression and spots physical distances shows a different and a non-linear behaviour in each dataset. Here we present Stardust*, a clustering algorithm which combines both transcriptomic and spatial information by using a non-linear formulation where the spatial contribution depends dynamically on the expression distances distribution in the surrounding space.

Methods

Stardust* is based on the clustering method implemented in the Seurat R package [5], a widely employed software for the analysis of single-cell RNA sequencing data.
Seurat’s main clustering method implements the Louvain algorithm [6] which represents each spot of a dataset as a node in a graph where the link between pairs of nodes encodes a predefined distance between gene expression profiles. This method is divided in two steps which respectively apply modularity optimization and community aggregation. Unlike Seurat, Stardust* replaces the distance matrix between transcriptional profiles given as input to the Louvain algorithm with a new distance matrix obtained by combining gene expression distances and spots spatial distances.
In particular, pairwise euclidean distances between gene expression profiles are calculated in the PCA space [7], while the euclidean distances between spots coordinates are weighted by the normalized values of the transcriptional distances distribution. The final distance matrix is given by the summation of these two vectors and it will be used by Seurat to find the neighbors in the network by computing the global distances or closeness between pairs of nodes representing the distance of each spot with respect to the others.
Clustering quality was assessed by computing three validation measures: the cell stability score, the coefficient of variation and the Moran’s index.
To compute the cell stability score we relied on the rCASC R package [8] that assigns to each spot a score representing how many times a spot remains clustered with the same set of elements after applying a selected number of perturbations to the dataset. A high-quality clustering result should not change when a perturbation is applied. The coefficient of variation is derived from the cell stability score distribution as the ratio between its standard deviation and mean. The lower the coefficient, the better the clustering. The Moran's I was used to evaluate the biological significance of the obtained clusters estimating the spatial autocorrelation of the genes.

Results

To test our clustering algorithm we chose five publicly available ST datasets provided by 10x Genomics (human lymph node, human heart, human breast cancer 1, human breast cancer 2 and mouse kidney), two public Seq-Scope datasets (colon and liver) and one dataset from the Slide-seqV2 technology (cerebellum). For each 10x dataset we computed three configurations of Stardust* with and without considering space, obtained by varying the clustering resolution, and we noticed that the introduction of spatial information improves clustering stability.
Comparing the results with other state-of-the-art tools for ST data clustering, namely BayesSpace [9], Giotto [10], stLearn [11] and spaGCN [12], we observed that Stardust* achieves higher or comparable stability scores with respect to the other compared methods.
Biological clustering validation was conducted by computing Moran’s index on 10x, Seq-Scope and Slide-seqV2 datasets and showing that genes with the highest spatial autocorrelation coefficients colocalize in clusters shapes achieved by Stardust*. Moreover, clustering structures were compared with the available pathologists annotations confirming the biological consistency of our method.

Info
Code repository: https://github.com/InfOmics/stardust References: [1] Noel, T. et al. Frontiers in Physiology 2317 (2022) [2] Moses, L. Nature methods 1-13 (2022). [3] Rodriques, Samuel G. et al. Science 363.6434: 1463-1467 (2019). [4] Cho, Chun-Seok et al. Cell 184.13: 3559-3572 (2021) [5] Butler, A. et al. Nature Biotechnology 36.5 (2018) [6] Blondel, Vincent D et al. Journal of Statistical Mechanics: Theory and Experiment 2008.10 (2008) [7] Jolliffe I.T. et al. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374.2065 (2016) [8] Alessandrì, L. et al. GigaScience 8.9 (2019) [9] Zhao, E. et al. Nature biotechnology 39.11: 1375-1384 (2021). [10] Dries, R. et al. Genome biology, 22(1), 1-31 (2021) [11] Pham, D. et al. BioRxiv (2020) [12] Hu, J. et al. Nature methods 18.11: 1342-1351 (2021)

filename	-
Figure	
-	
Availability	https://www.biorxiv.org/content/10.1101/2022.04.27.489655v1.full

Corresponding Author	
Name, Surname	Simone, Avesani
Email	simone.avesani@univr.it
Submitted on	10.05.2022