BITS :: Call for Abstracts 2021 - Oral communication

Type	Oral communication
Session	Multiomics and Single Cell Analysis
Title	Copy Number Variation inference from single-cell expression data with the R package scMuffin.
All Authors	Valentina Nale [1], Alice Chiodi [1], Ingrid Cifola [1], Cinzia Cocola [1], Noemi Di Nanni [1], Eleonora Piscitelli [1], Rolland Reinbold [1], Luciano Milanesi [1], Alessandra Mezzelani [1], Paride Pelucchi [1]*, Ettore Mosca [1]*

Affiliation

[1] CNR-ITB, Institute of biomedical technologies

Motivation

Single cell (SC) analysis is crucial to study the complex cellular heterogeneity of solid tumors, which is one of the main obstacles for the development of effective cancer treatments. Such tumors typically contain a mixture of cells with aberrant genomic and expression profiles affecting specific sub-populations that have a pivotal role in cancer progression, whose identification eludes bulk approaches. We are developing a MUlti-Features INtegrative approach for SC data analysis (scMuffin) [1] that characterizes cell identity on the basis of multiple and complementary criteria, including the Copy Number Variation (CNV) inference from SC expression data. Here, we present CNV estimation method in our scMuffin package, the comparison with other CNV inference tools and the results obtained analyzing a public dataset of human gliomas [6].

Methods

scMuffin is implemented in R language [1]. CNV inference in scMuffin package is based on the “adjacent gene windows” approach described in [2] and [3]. Briefly, the CNV profile of each single cell is calculated by applying a moving average over scaled gene expression levels ordered by genomic location, and by subtracting a normal reference profile. The expression profiles of normal cells were collected from GTEx [7] or found within the sample on the basis of the single-cell “CNV signal”, which summarizes the CNV profile of each cell in a unique value. The most computationally expensive steps are implemented in parallel.

Results

The workflow of CNV estimation in scMuffin provides the following functions: i) SC data filtering; ii) calculation of the genomic windows-by-cells matrix; iii) comparison of CNV profiles with reference profiles; iv) calculation of the single-cell CNV signal; v) CNV visualization. We run our workflow as well as two other tools (InferCNV and CopyKat) [4,5] on public SC expression data of human gliomas [6]. The CNV profiles obtained with scMuffin determine a series of clusters that are characterized by the absence or presence of relevant deviations of the CNV score in one or more genomic locations, indicating normal or aberrant cells respectively. This pattern is coherent with that reported in the study on human gliomas [6], where a cell malignancy score was calculated on the basis of a Principal Components Analysis of cell’s chromosome-level gene expression averages. As expected, CNV estimation by scMuffin leads to results that are more similar to those from InferCNV than from CopyKat, since scMuffin and InferCNV share similar estimation principles. CNV estimation in scMuffin is faster than in the other two tools and, importantly, CNV profiles and CNV clustering can be integrated with the results of other analyses provided by scMuffin, such as cluster annotation using known markers expression, comparison between CNV clusters and global expression profile clusters, or comparison of CNV state and cell state trajectory analysis. CNV inference from SC expression data in cancer provides the opportunity of distinguishing normal and malignant cells and reconstructing a clonal substructure. However, our comparison of CNV inference tools underlined a series of open issues, like the presence of missing values and the sensitivity of the results to the considered reference, which encourage further developments in this field.

Info

References:
[1] N. Di Nanni et al., scMuffin: an R package for resolving solid tumor heterogeneity from single-cell data, ESHG. June 12-15, 2021 (accepted)
[2] Patel A. P., et al., Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science (2014)
[3] Tirosh, I. et al., Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science (2016)
[4] Tickle, T. et al., Infer Copy Number Variation from Single-Cell RNA-Seq Data. Bioconductor (2019).

inferCNV of the Trinity CTAT Project https://github.com/broadinstitute/inferCNV
[5] Gao, R. et al., Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. Nature Biotechnology (2021)
[6] Yuan, J. et al., Single-cell transcriptome analysis of lineage diversity in high-grade glioma. Genome Medicine (2018)
[7] Aguet, F. et al., GTEx Consortium atlas of genetic regulatory effects across human tissue. Science (2020)

Figure

-

Availability -

Corresponding Author

Name, Surname Valentina, Nale

Email valentina.nale@itb.cnr.it

Submitted on 07.05.2021