POLITECNICO DI TORINO
Repository ISTITUZIONALE

Single-cell DNA Sequencing Data: a Pipeline for Multi-Sample Analysis

(Article begins on next page)

04 August 2020

heatmap, for each sample: in this way, it is possible to compare the results of the two tools, also, in a qualitative way. Moreover, in order to provide a functional information, CNV events are annotated with the corresponding gene symbols. Finally, density plots highlight, for each sample, the clusters of cells with a similar mean ploidy, allowing to set a threshold and filter out cells which ploidy is considered irrilevant for the following analysis.

Multi-sample analysis. Multi-sample data are, finally, aggregated to evaluate evolutionary distance among them. This task is accomplished by another module which performs the phylogenetic reconstruction algorithm on all cells, from all different samples, and produces one single tree and one single heatmap. In order to evaluate intra and inter-tumor heterogeneity, two distance metrics have been defined: (i) an intra-tumor (intra-TH) heterogeneity measure, which tells how distant cells within the same sample are, and (ii) an inter-tumor heterogeneity (inter-TH) measure, which tells how distant cells from different sample are. Both measures are based on the distances computed to construct the phylogenetic tree and associated to its branches.

3 CONCLUSION AND FUTURE WORK

We have developed a platform for large-scale single-cell DNA CNV analysis, capable of enriching the results of a proprietary scDNA pipeline, Cell Ranger DNA, and exploit them to evaluate intra and inter-tumor heterogeneity, with the help of a flexible and open source tool, Ginkgo. Additionally, we plan to implement a new module capable of performing cell clustering starting from their CNV profiles: this should allow to better investigate spatial segregation of clones. In conclusion, this pipeline is a ready-to-use instrument for researchers who are interested in exploring temporal and spatial tumor evolution and need automatic tools to help them managing data and performing analyses.

REFERENCES

[1] 10x Genomics. [n. d.]. 10x Genomics: Biology at True Resolution. Retrieved April 15,2 2019 from https://www.10xgenomics.com/
[2] Tyler Garvin, Robert Aboukhalil, Jude Kendall, Timour Baslan, Gurinder S Atwal, James Hicks, Michael Wigler, and Michael C Schatz. 2015. Interactive analysis and assessment of single-cell copy-number variations. *Nature Methods* 12 (Sept. 2015), 1058–1060. https://doi.org/10.1038/nmeth.3578