Evaluating the Reproducibility of Single-Cell Gene Regulatory Network Inference Algorithms

Yoonjee Kang[1], Denis Thieffry[1], Laura Cantini[1]

[1]*Computational Systems Biology Team, Institut de Biologie de l'Ecole Normale Supérieure, CNRS UMR 8197, INSERM U1024, Ecole Normale Supérieure, Paris Sciences et Lettres Research University, Paris, France*

Networks are powerful tools to represent and investigate biological systems [1]. The development of algorithms inferring regulatory interactions from functional genomics data has been an active area of research [1,2]. With the advent of single-cell RNA-seq data (scRNA-seq), numerous methods specifically designed to take advantage of single-cell datasets have been proposed [3,4]. However, published benchmarks on single-cell network inference are mostly based on simulated data [3,4]. Once applied to real data, these benchmarks take into account only a small set of genes and only compare the inferred networks with an imposed ground-truth. Here, we benchmark six single-cell network inference methods based on their reproducibility, i.e., their ability to infer similar networks when applied to two independent datasets for the same biological condition. We tested each of these methods on real data from three biological conditions: human retina, T-cells in colorectal cancer, and human hematopoiesis. Once taking into account networks with up to 100,000 links, GENIE3 results to be the most reproducible algorithm and, together with GRNBoost2, show higher intersection with ground-truth biological interactions. These results are independent from the single-cell sequencing platform, the cell type annotation system and the number of cells constituting the dataset. Finally, GRNBoost2 and CLR show more reproducible performance once a more stringent thresholding is applied to the networks (1,000–100 links). In order to ensure the reproducibility and ease extensions of this benchmark study, we implemented all the analyses in scNET, a Jupyter notebook available at https://github.com/ComputationalSystemsBiology/scNET.

References and useful links

1. Barabási, A.-L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. doi: 10.1038/nrg1272

2. The DREAM5 Consortium, Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., et al. (2012). Wisdom of crowds for robust gene network inference. Nat. Methods 9, 796–804. doi: 10.1038/nmeth.2016

3. Chen, S., and Mar, J. C. (2018). Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. BMC Bioinformatics 19:232. doi: 10.1186/s12859-018-2217-z

4. Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A., and Murali, T. M. (2020). Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. Nat. Methods 17, 147–154. doi: 10.1038/s41592-019-0690-6