Robustness and applicability of functional genomic tools on scRNA-seq data

Christian H. Holland1,2, Jovan Tanevski1,3, Jan Gleixner4,5, Manu P. Kumari6, Elisabetta Mereu7, Brian A. Joughin8,9, Oliver Stegle4,5, Douglas A. Lauffenburger6, Holger Heyn7,10, Bence Szalai11 and Julio Saez-Rodriguez1,2,9

1 Institute for Computational Biomedicine, Heidelberg University, Faculty of Medicine, Bioquant – Im Neuenheimer Feld 267, 69120 Heidelberg, Germany. 2 Joint Research Centre for Computational Biomedicine (JRC-COMBINE), RWTH Aachen University, Faculty of Medicine, Pauwelsstrasse 19, 52074 Aachen, Germany. 3 Department of Knowledge Technologies, Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia. 4 German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. 5 European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Meyerhofstrasse 1, 69117 Heidelberg, Germany. 6 Department of Biological Engineering, MIT, Cambridge MA. 7 CNAG-CRSI, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. 8 Koch Institute for Integrative Cancer Biology, MIT. 9 European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, CB10 1SD, Histon, Cambridge. 10 Universitat Pompeu Fabra (UPF), Barcelona, Spain. 11 Semmelweis University, Faculty of Medicine, Department of Physiology, Budapest, Hungary. * To whom correspondence should be addressed: julio.saez@bioquant.uni-heidelberg.de

Introduction

- Single-cell RNA sequencing (scRNA-seq) data has specific characteristics and limitations such as drop-out events.
- It is not clear if functional genomic tools established for bulk sequencing can be applied on scRNA-seq.
- We perform benchmark studies on *in silico* and *in vitro* scRNA-seq data.
- We focus on the tools VIPER (using DoRothEA) and PROGENy that estimate transcription factor (TF) and pathway activities, respectively.
- We also test the performance of VIPER and PROGENy in a more heterogeneous system that would illustrate a typical scRNA-seq data analysis scenario.

Benchmark on Simulated Data

Case Study on Human Cell Atlas Data

- B cells - HEK cells - Monocytes + Dendritic + T + NK cells

Conclusions & Outlook

- Our systematic and comprehensive benchmark study suggests that VIPER (using DoRothEA as gene regulatory network) and PROGENy can functionally characterise scRNA-seq data.
- We provide recommendations on how to use DoRothEA and PROGENy dependent on various scRNA-seq protocols.
- We will extend this analysis to other tools such as AUCell, metaVIPER and pathway analysis with GO gene sets.
- The best performing tool will be used to decipher key molecular mechanisms of chronic liver diseases on single-cell and bulk level.

Funding & Contact

**LiSyM**

Federal Ministry of Education and Research

Questions? Contact me: christian.holland@bioquant.uni-heidelberg.de

www.lisym.org

www.saezlab.org