BITS :: Call for Abstracts 2021 - Oral communication

Type	Oral communication
Session	Systems Biology
Title	Single-Cell reaction expression analysis as a bridge between transcriptomics and fluxomics
All Authors	Galuzzi B.G (1), Damiani C(1)
Affiliation	

(1) Department of Biotechnology and Biosciences, University of Milan-Bicocca, Milan

Motivation

Several studies have successfully demonstrated how Constrained-Based Reconstruction and Analysis (COBRA) methods can be used to perform metabolic analysis through the integration of omics data into a genome-scale metabolic network. For example, Flux Balance Analysis uses linear programming to calculate a feasible flux distribution that maximizes the flux through an objective function.
Single-cell RNA sequencing (scRNA-seq) goes one step further with respect to bulk one. Indeed, it allows studying gene expression at an unprecedented resolution. However, while the integration of bulk transcriptomics data into a metabolic network has been already explored, few works [Damiani et al., 2019] have studied the integration of scRNA-seq data. scRNA-seq data require special attention as they are more subject to noise and false negatives.
The classical way to integrate transcriptional data is through the Gene-Protein-Reaction association rules (GPRs), which are logical formulas describing how gene products concur to catalyze a given reaction, using boolean operators. In this sense, the GPRs represents a bridge between transcriptomics (i.e. the expression of genes involved in the metabolism) and fluxomics, even if the presence of false-negative in single-cell RNA-seq data could affect the GPR computation. In this work, we investigate the best practices to integrate scRNA-seq data via GPRs.

Methods

We used Reaction Activity Scores [Graudenzi et al., 2018] to map RNA-seq data (i.e., the gene count matrix) into a new reaction expression data, which consists of a NxM matrix, where N is the number of reactions involved in the metabolic model and M is the number of single-cells under study.
Starting from the RNA-seq count matrix and a list of metabolic reactions, we computed the RAS matrix, substituting the mRNA abundances in the corresponding GPRs. We tried different possible operations to evaluate the logical operators of the GPRs: sum and maximization for the OR operator and mean and minimization for the AND operator.
Finally, following current best practices in single-cell RNA seq analysis [Luecken and Theis, 2019], we performed clustering and differential reaction expression (DRE) analysis on the RAS matrix to determine which reactions are differentially “expressed” between different conditions. For clustering, we used the Leiden algorithm after applying PCA for dimensional reduction of the data. Then, we examined differentially expressed features (i.e. reactions) and identified positive and negative markers of a single cluster, compared to all other cells.

Results

We used a scRNA-seq dataset from the NCBI GEO data repository GSE110949 [Banerjee et al., 2019]. This dataset is obtained from MDA-MB-231 breast cancer cells in high glucose media, treated with 2mM Metformin for either a short or long period, and then cultured in mammospheres. In this dataset, 14698 genes are considered across 651 cells.

We integrated the data into a metabolic core model (422 metabolites and 495 reactions) and computed the RAS matrix for all the reactions provided with a GPR.

Since we have prior information about the cells (i.e. adapted or not adapted to Metformin), we included such information as ground-truth information to perform cluster analysis, selecting the best hyper-parameters for clustering (PCA components, number of neighbours, resolution) as the ones which maximize the v-measure, which represents one of the most common metrics to measure the goodness of clustering.

Figure 1 summarizes the preliminary results of the differential expression reaction (DER) analysis. The max and min operator results in the best methods for evaluating AND or OR operator. Cluster analysis partitioned cells into four clusters. Most notable is the presence of cluster 3 formed by non adapted cells mainly, whereas there is no significant preponderance of one group or another into the other three clusters. DRE

analysis finds highly differential reactions in each cluster.

Info

References

Banerjee, A., Birts, C. N., Darley, M., Parker, R., Mirnezami, A. H., West, J., ... & Blaydes, J. P. (2019). Stem cell-like breast cancer cells with acquired resistance to metformin are sensitive to inhibitors of NADH-dependent CtBP dimerization. Carcinogenesis, 40(7), 871-882.
Damiani, C., Maspero, D., Di Filippo, M., Colombo, R., Pescini, D., Graudenzi, A., ... & Mauri, G. (2019). Integration of single-cell RNA-seq data into population models to characterize cancer metabolism. PLoS computational biology, 15(2), e1006733.
Graudenzi, A., Maspero, D., Di Filippo, M., Gnugnoli, M., Isella, C., Mauri, G., ... & Damiani, C. (2018). Integration of transcriptomic data and metabolic networks in cancer samples reveals highly significant prognostic power. Journal of biomedical informatics, 87, 37-49.
Luecken, M. D., & Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. Molecular systems biology, 15(6), e8746.

Figure

Availability	https://drive.google.com/file/d/14ez-68bZRfKLX3BNcH7hxR-HtFmaWjnz/view?usp=sharing

Corresponding Author

Name, Surname	Bruno Giovanni, Galuzzi
Email	bruno.galuzzi@unimib.it

Submitted on	09.05.2021