Functional module detection through integration of single-cell RNA-sequencing data with protein–protein interaction networks

Florian Klimm[1], Enrique M. Toledo[2], Thomas Monfeuga[2], Fang Zhang[2], Charlotte M. Deane[1], and Gesine Reinert[1]

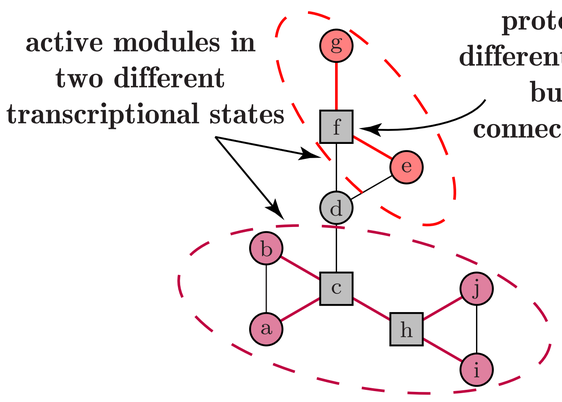
[1]Department of Statistics, University of Oxford, [2]Discovery Technology and Genomics, Novo Nordisk Research Centre Oxford

bioRxiv: https://doi.org/10.1101/698647

Summary

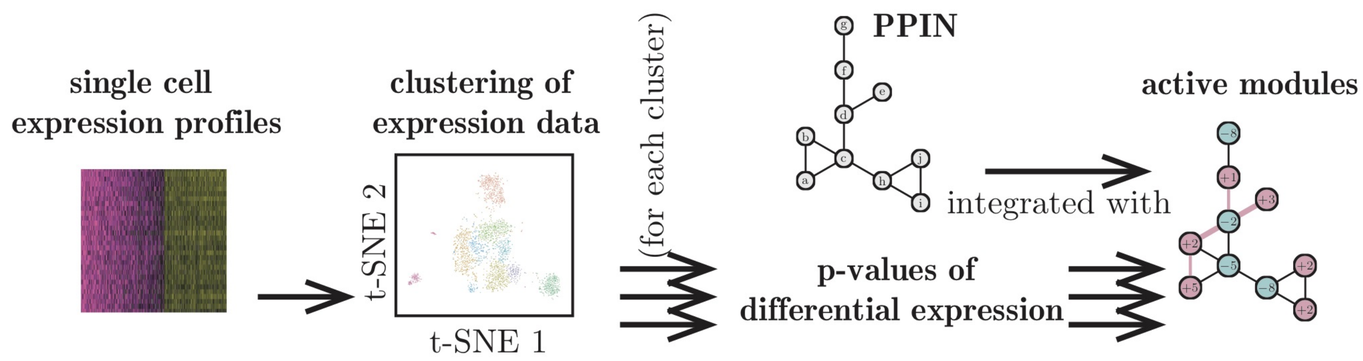
Protein–protein interaction networks (PPINs) are abstract representations of interactions between proteins. While there has been considerable insight from investigating PPINs alone [1], analysing them in an integrated way, together with gene-expression data provides biological context [3]. In recent years, much attention has been given to single-cell RNA-sequencing (scRNA-seq) techniques as they allow researchers to study and characterise tissues at single-cell resolution [4]. We present scPPIN, a method to integrate such scRNA-seq data with PPINs to detect active modules in cells of different transcriptional states [5]. As a case study, we investigate scRNA-seq from human liver spheroids but these techniques are also applicable to other organisms and tissues. This novel method allows us to identify proteins important in liver metabolism that could not be detected from gene expression data alone. Furthermore, we can associate cells in a given transcriptional state with enriched biological pathways.

Benefits of this method



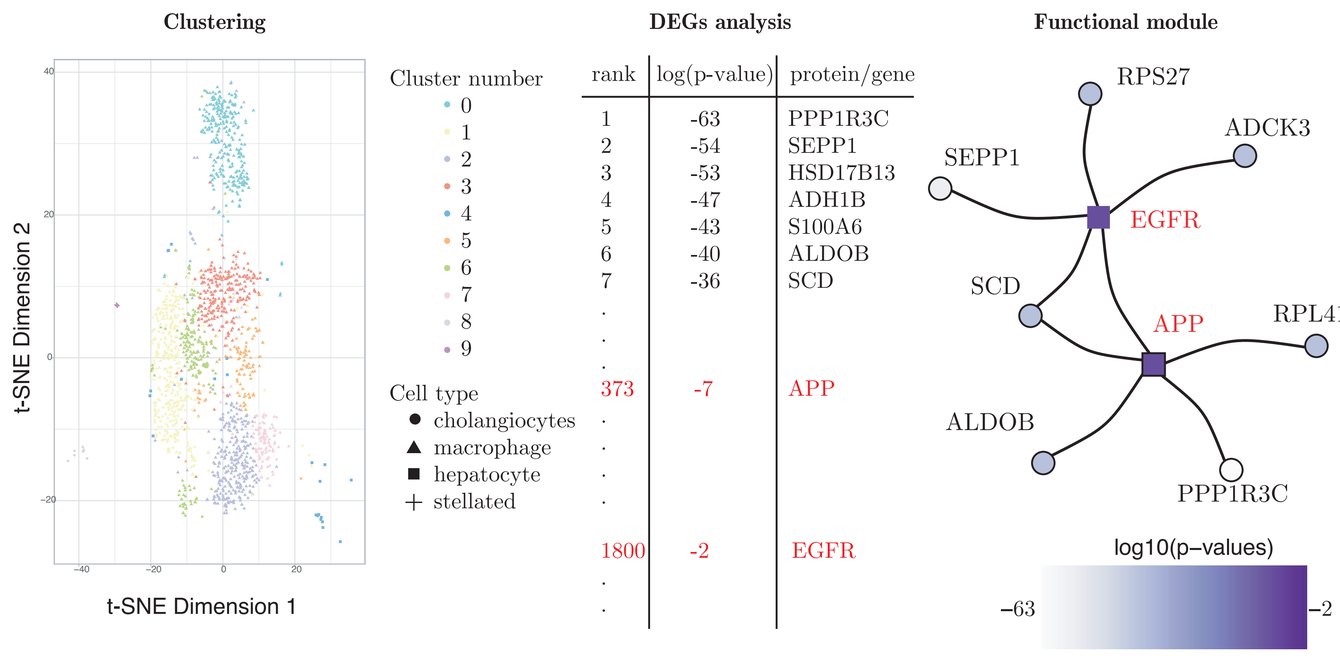
1. We associate each cell cluster (which represent cell types or transcriptional states) with an active module in the PPIN.
2. This identifies proteins that are *not* differentially expressed and thus would have been missed by a differential expression analysis alone.

Method



1. Clustering of scRNA-seq data (e.g., with Seurat [2]).
2. Computation of p-values of differential expression for each cluster.
3. Estimation of node scores by fitting a *beta-uniform model* with a maximum-likelihood approach [3].
4. Combination of these node scores with a reference PPIN to construct node-weighted PPINs for each cluster.
5. Computation of functional modules with maximal change in gene expression as maximum-weight connected subgraphs with dapcstp [6].

Liver Data Application



As a case study, we measure scRNA-seq in human primary hepatocytes. We use Seurat to cluster the cells and to identify differentially expressed genes (DEG). We construct a PPIN from the publicly available BioGrid database, version 3.5.166. We use scPPIN to detect functional modules for each cluster of hepatocytes. Here, we show one example: A functional module consisting of nine proteins. This includes the proteins expressed by the genes that have the seven lowest p-values (e..g, PPP1R3C) but also two additional proteins: APP and EGFR, which are both membrane-bound receptors.

If we use scPPIN to detect functional modules in different cell clusters, we obtain different functional modules that are associated with different biological functions in the liver, e.g., 'response to inflammation' or 'alcohol metabolic process'.

References

[1] Waqar Ali, Charlotte M. Deane, and Gesine Reinert. Protein interaction networks and their statistical analysis. In Michael P. H. Stumpf, David J. Balding, and Mark Girolami, editors, *Handbook of Statistical Systems Biology*, pages 200–234. John Wiley & Sons, Ltd Chichester, UK, 2011.

[2] Andrew Butler, Paul Hoffman, Peter Smibert, Efthymia Papalexi, and Rahul Satija. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*, 36(5):411, 2018.

[3] Marcus T Dittrich, Gunnar W Klau, Andreas Rosenwald, Thomas Dandekar, and Tobias Müller. Identifying functional modules in protein–protein interaction networks: an integrated exact approach. *Bioinformatics*, 24(13):i223–i231, 2008.

[4] Byungjin Hwang, Ji Hyun Lee, and Duhee Bang. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental & Molecular Medicine*, 50(8):96, 2018.

[5] Florian Klimm, Enrique M Toledo, Thomas Monfeuga, Fang Zhang, Charlotte M Deane, and Gesine Reinert. Functional module detection through integration of single-cell RNA sequencing data with protein-protein interaction networks. *bioRxiv*, 2019.

[6] Markus Leitner, Ivana Ljubić, Martin Luipersbeck, and Markus Sinnl. A dual ascent-based branch-and-bound framework for the prize-collecting steiner tree and related problems. *INFORMS Journal on Computing*, 30(2):402–420, 2018.

Availability

The scPPIN method is available as an R library and as an online-tool under

`github.com/floklimm/scPPIN`

Acknowledgements

This project is funded by a Novo Nordisk–University of Oxford pump priming award under the Strategic Alliance between the two partners and the EPSRC.