SingleCellHaystack: A clustering-independent method for finding differentially expressed genes in single-cell transcriptome data

Alexis Vandenbon[1,2] and Diego Diez[3]

[1] Institute for Frontier Life and Medical Sciences, Kyoto University, 53 Shougoin Kawara-cho, Sakyou-ku, Kyoto 606-8507, Japan

[2] Institute for Liberal Arts and Sciences, Kyoto University, Yoshidanihonmatsu-cho, Sakyo-ku, Kyoto 606-8501, Japan

[3] Immunology Frontier Research Center, Osaka University, 3-1 Yamada-oka, Suita, Osaka 565-0871, Japan

Single-cell data has the potential to reveal new insights into the origins and progression of diseases, including the prediction of new marker genes and regulatory interactions. A common analysis of single-cell data includes clustering of cells and identifying differentially expressed genes (DEGs). How cell clusters are defined has important consequences for downstream analyses and the interpretation of results, but is often not straightforward. To address this difficulty, we present singleCellHaystack, a method that enables the prediction of DEGs without relying on explicit clustering of cells. Our method uses Kullback–Leibler divergence to find genes that are expressed in subsets of cells that are non-randomly positioned in a multidimensional space. Comparisons with existing DEG prediction approaches on artificial datasets show that singleCellHaystack has higher accuracy. We illustrate the usage of singleCellHaystack through applications on 136 real transcriptome datasets and a spatial transcriptomics dataset. We demonstrate that our method is a fast and accurate approach for DEG prediction in single-cell data. singleCellHaystack is implemented as an R package and is available from CRAN and GitHub.

Reference:

Vandenbon A. and Diez D. A clustering-independent method for finding differentially expressed genes in single-cell transcriptome data, *Nature Communications*, 11 (1), 4318, 2020.