Unsupervised feature selection for manifold alignment of scRNA-seq data

Yutong Wang 1, Tasha Thong2, Justin Colacino2, Venkatesh Saligrama3, Laura Balzano1, Clayton Scott1,

1Department of EECS, 2Department of Environmental Health Sciences, University of Michigan, Ann Arbor, MI; 3Department of ECE, Boston University, Boston, MA

Abstract

We introduce a novel unsupervised feature selection method for manifold alignment of single-cell RNA-seq datasets. We demonstrate that our gene selection method, CorGI (Correlation matrix singular value Gap Inflation), filters out much of the batch effects.

Measuring batch effects in gene sets

Figure 1. * HDG = Highly Dropped-out Genes [3]. Batch mixing \( \propto \) batch label prediction error

Figure 2: (A) Input to CorGI. (B) Main loop. (C) Selection of top genes.

Acknowledgement

Figure 3. Pre-implantation embryo development [1, 5]. Format of titles of scatterplots: Gene set name, batch label prediction error

Figure 4. Neurogenesis in the subventricular zone [2, 4]

For more details, please visit web.eecs.umich.edu/~yutongw

Code and preprint will be available soon.

References