ACE: Explaining single-cell cluster from an adversarial perspective

Yang Lu[1], Timothy C. Yu[2], Giancarlo Bonora[1], William S. Noble[1,3]

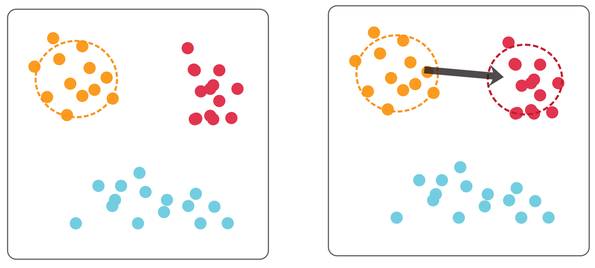
1. Department of Genome Sciences, University of Washington, Seattle, WA
2. Graduate Program in Molecular and Cellular Biology, University of Washington, Seattle, WA
3. Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA

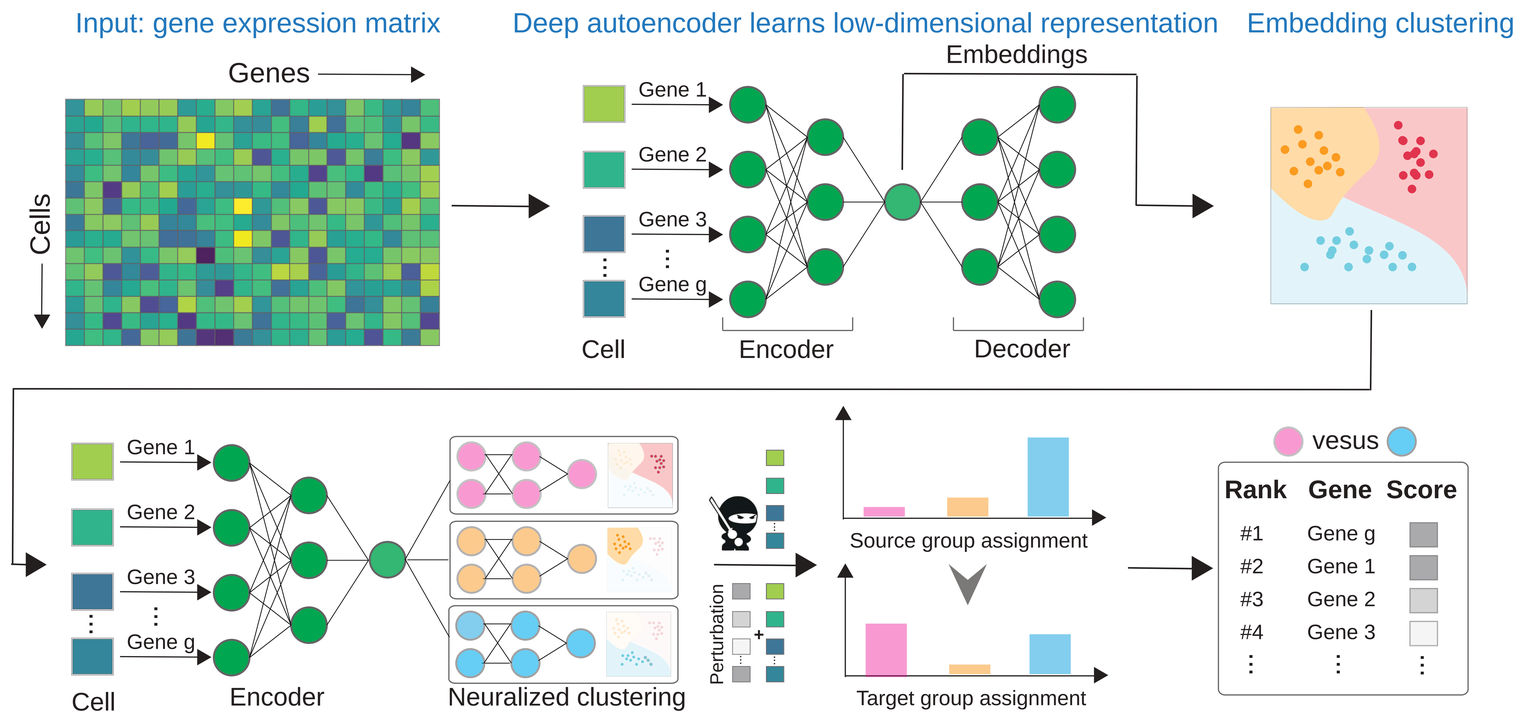
Overview

ACE (Adversarial Clustering Explanation) aims to identify differential genes in scRNA-seq data by joingly modelling:

- The uncertainty induced in nonlinear embedding of scRNA-seq data.
- The stochasticity in cluster assignments.
- The dependency among genes.
- Both enriched and depleted genes relative to other cell types.

ACE supports one-vs-one and one-vs-rest explanation:



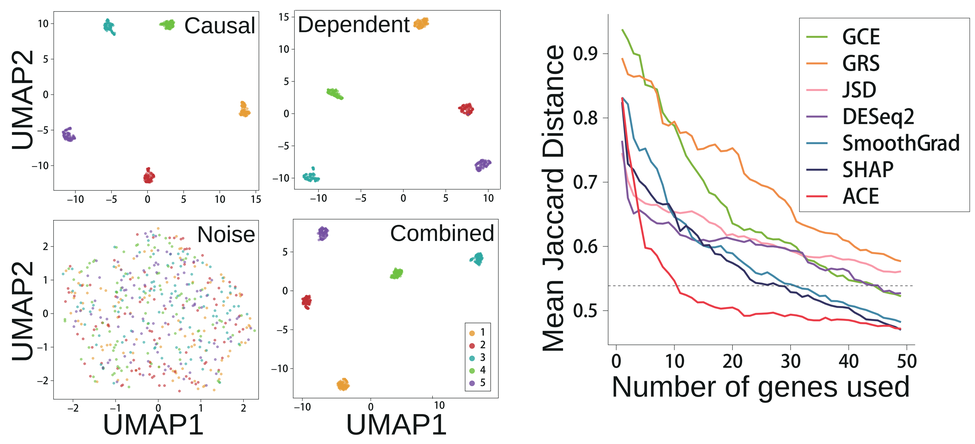


Performance on simulated data

We use SymSim [1] to simulate causal and noise genes for 500 cells

- 20 causal genes, 100 noise genes
- 100 dependent genes are weighted sums of random causal genes with noises

We use Mean Jaccard Distance to quantify how well the top k genes in the ranking capture the clustering structure

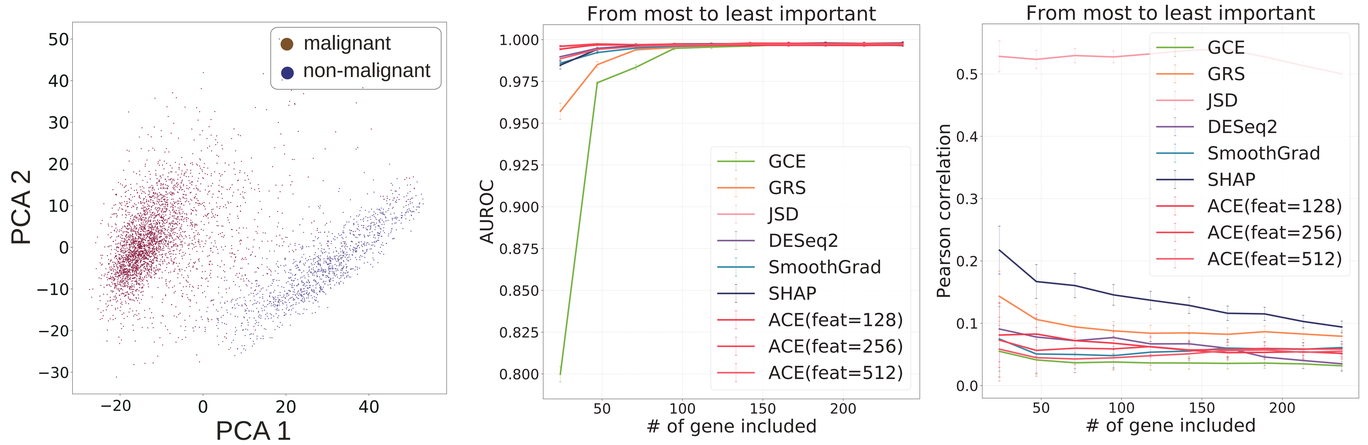


Performance on melanoma dataset [2]

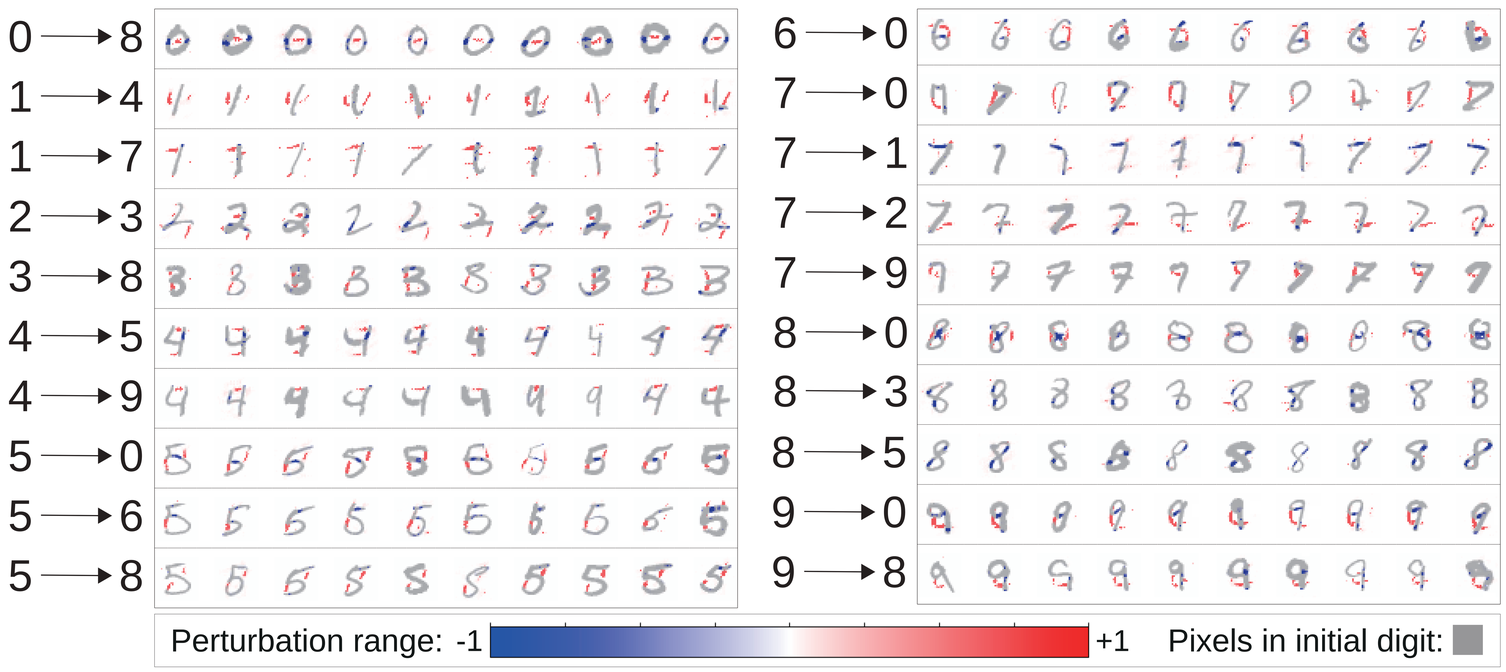
The cell labels are determined by copy number variation instead of data-driven

- Two cell types: malignant vs. non-malignant
- 4513 cells (1257 malignant and 3256 non-malignant), 23686 genes

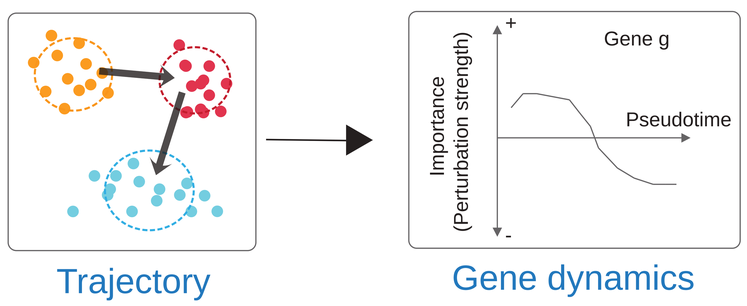
We want the top k genes in the ranking to be both highly discriminative and non-redundant



Performance on MNIST dataset



Future work



Reference

[1] Zhang, et al. "Simulating multiple faceted variability in single cell RNA sequencing." Nature Communications 10.1 (2019): 1-16.

[2] Tirosh, et al. "Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq." Science 352.6282 (2016): 189-196.