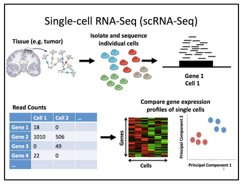
Interactive Machine Learning for Gene Module Discovery in Transcriptomics

Eli T. Brown

Laboratory for Interactive Human-Computer Analytics
lihca.io

DePaul University College of Computing and Digital Media

Enter: Single-Cell mRNA Sequencing



scRNAseq allows full snapshot of gene expression…

But it produces large noisy data

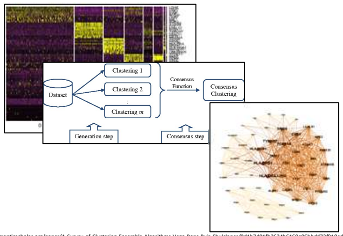
Machine learning to the rescue?

Our Approach

Perform clustering with multiple parameter sets

Merge into a distance metric for visualization

We think we can "make life better for thousands of people" -collaborator



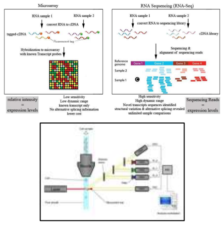
https://www.semanticscholar.org/paper/A-Survey-of-Clustering-Ensemble-Algorithms-Vega-Pons-Ruiz-Shulcloper/0d1b7d01fb2634b6160a96bbdd73f918ed3859cb

BioTech Problem

Collaborator at MIT is studying *macrophages* (human immune cells)

One subject: exposure to TB affects people differently

To learn why must study the genes, but typical approaches fall short



https://www.otogenetics.com/rna-sequencing-vs-microarray/
https://learn.gencore.bio.nyu.edu/single-cell-rnaseq/

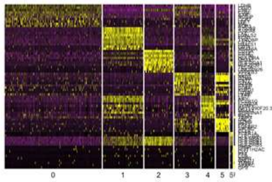
BioTech: Single-Cell mRNA Sequencing

Biologists use (bi)clustering algorithms that are very sensitive to parameters

Tedious, technical process

Inspect by hand…And repeat…

Only human operator can determine what is actually interesting.



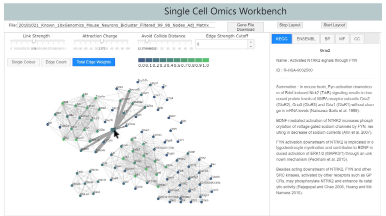
Transcriptomics Workbench

Uses ensemble clustering to create layout of genes

Interaction with layout for group discovery

Gene database information at fingertips

Future: rule-based IML



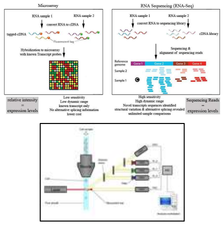
Supported by CDM Summer Research 2019

BioTech Problem

Typical approaches:

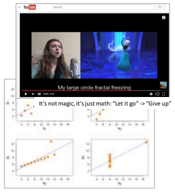
Microarrays, flow cytometry

Limited to testing specific proteins in an experiment



https://www.otogenetics.com/rna-sequencing-vs-microarray/
https://learn.gencore.bio.nyu.edu/single-cell-rnaseq/

Motivating Assertion in My Work



Human + Machine
>
Human XOR Machine