Final Project - MultiK

true   true

Contents

Background Info

Goal

This is for INFO 523 - final project. I'm going to **attempt** to implement the multiK package.

MultiK

MultiK is meant to **guide** your clustering efforts for scRNA-seq data. If I remember, I'll link the Youtube video of me explaining wtf I'm talking about.

Please note

There is no **ideal** clustering. There's not necessarily a *right* answer. There's information to be had and debate for any amount of clustering with this type of data. Therefore, this is one way of exploring clustering and it can inform the final clustering decisions.

Programming

0. Always figure out your working directory/path situation

```
library(here)
```

```
## here() starts at /Users/banka/Docs/ua_yr2_f22_sp23/final_project_multik
```

1. Load in your packages

Worked Example

1. MultiK comes with a dataset we can run a trial on

```
# it's a mixture of 3 cell lines (types)
example_data <- MultiK::p3cl
example_data
```

```
## An object of class Seurat
## 21247 features across 2609 samples within 1 assay
## Active assay: RNA (21247 features, 0 variable features)
```

2. Run the main algo to determine an appropriate # of clusters

```
# I've set this not to run when it knit because it takes 3 hours to run
# multik <- MultiK::MultiK(example_data, reps = 100)
```

3. Let's make those diagnostic plots

```
#MultiK::DiagMultiKPlot(multik$k, multik$consensus)
```

Let's look at those results!

```
#clusters <- getClusters(example_data, 3)
```

```
#pval <- CalcSigClust(example_data, clusters$clusters)
```

```
#PlotSigClust(example_data, clusters$clusters, pval)
```

And then you can add that info back onto your Seurat & use it as you see fit for downstream analyses.

I've chosen not to include this because it's outside the scope of this project!

Thanks so much!

You can find the presentation at