Upload Application Steps

Contents

a.phateR

Due to issues with different packages, PhateR can have issues loading. Please test phateR before running the application, using the following code:

```
library(reticulate)
Sys.setenv(RETICULATE_PYTHON = "/software/anaconda3/bin/python")
use_python("/software/anaconda3/bin/python")
reticulate::py_discover_config(required_module="phate")
reticulate::py_config()


library(phateR)
tree.phate <- phate(tree.data$data)
```

Please note, you may need to remove all objects from the workspace and quit the current R session for the code to work.

b.Change wd

The working directory should be changed in the code to load/save files (setwd()). It can be done on line 4 of app.R

```
1
2  ##PLEASE CHANGE THE WORKING DIRECTORY###
3
4  setwd("/data/2623287c/Project1/upload_app")
5
6  source("tcell_libs_raw_dash.R", local = TRUE)
```

c. UniProt file

The UniProt file should be ordered with the gene names first and the UniProt links second. The column names can then be selected in the app for the gene names and UniProt links.

d. Integration

As the workflow uses Seurat integration, it requires Seurat objects from more than one group

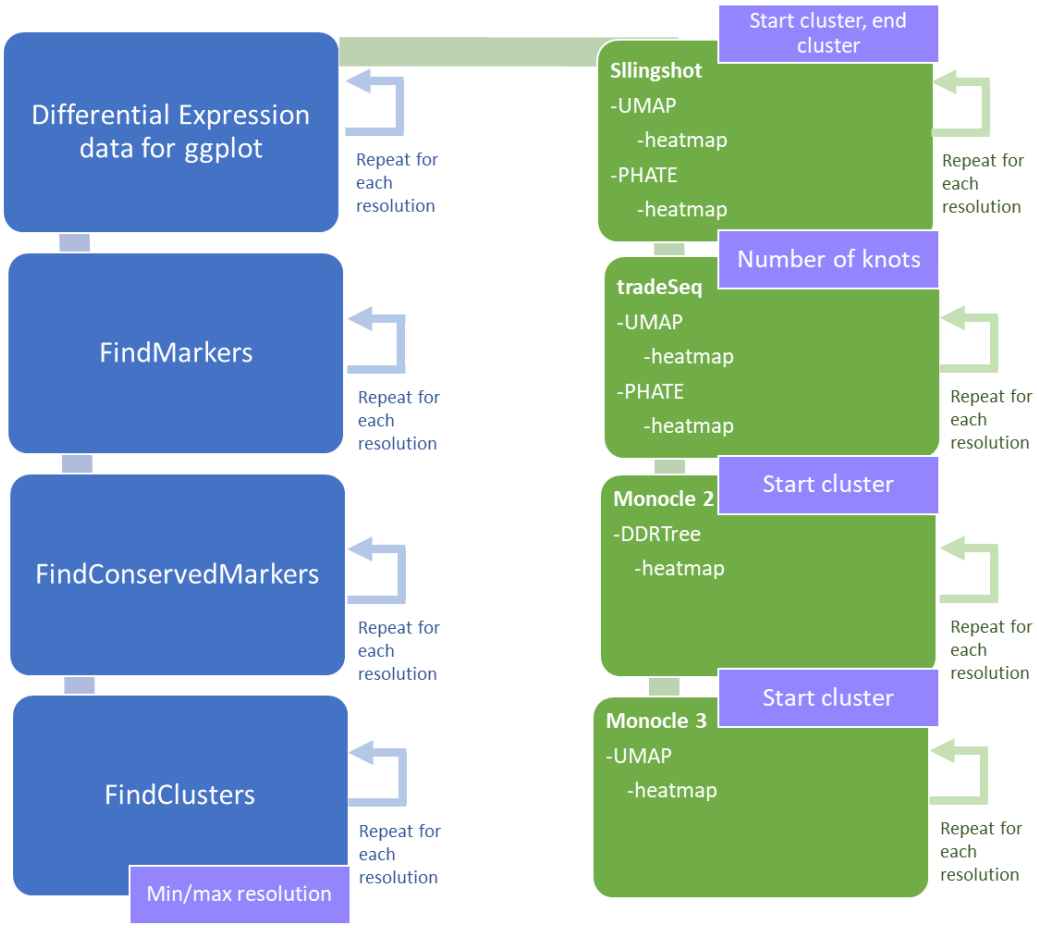
e.Pseudotime Parameters

Resolution - Choose resolutions from 0.15-.55. One resolution can be chosen if you already know the resolution best for the data, to prevent excess time running the code.

Table xx describes the parameters.

Tool	Parameter	Description	Default
Slingshot	Start cluster	Root cluster(s) for pseudotime analysis	NULL
	End cluster	End cluster(s) for pseudotime analysis	NULL
tradeSeq	Number of knots	Used in fitGAM. Recommended 3-10	3
Monocle 2	Root state	Monocle 2 is not cluster based so it finds the state associated with that cluster to use as the root state	NULL
Monocle 3	Root node	Monocle 3 is not cluster based so it finds the cells associated with that cluster to use as the root node	NULL

f. Overview of Upload Pseudotime Analysis



Single Cell Analysis

Single cell analysis performed according to Seurat integration vignette (https://satijalab.org/seurat/articles/integration_introduction.html). FindClusters, FindConservedMarkers, FindMarkers and Differential Expression data for ggplot are repeated for each resolution, as chosen by the user.

PHATE reduction is added to each resolution.

Psuedotime Analysis

Slingshot

For each resolution, the Seurat object is converted to single cell experiment and the dimensionality reduced. Slingshot is then run using the Seurat cluster labels with the start and end cluster(s) chosen by the user. It is run for both UMAP and PHATE. The outputs from Slingshot are converted to slingshot datasets.

heatmaps

For each resolution, the Seurat package is used to discern Variable Features. The p/q-value for these variable genes are calculated using gam from the package gam. The results are sorted by q-

value. The heatmaps are annotated with clusters and lineages. Heatmaps are produced for UMAP and PHATE

tradeSeq

For each resolution, the clusters and counts are extracted from the single cell Slingshot output and Seurat object, respectively. The counts are filtered to remove those with a count >1 in >=120 cells. This significantly speeds up the code and was recommended by the developer. Next, fitGAM is used to produce the final output of tradeSeq with the knots parameter set to the value chosen by the user.

Heatmaps

For each resolution, associationTest is used to assess whether gene expression is associated with pseudotime. They are sorted by P-value (due issues with q-value in the tradeSeq package). The heatmaps are annotated with clusters and lineages. Heatmaps are produced for UMAP and PHATE.

Monocle 2

For each resolution, count, phenotype and feature data are extracted from the Seurat object to produce a cell data set. Next, standard Monocle 2 pipeline is carried out (estimateSizeFactors, estimateDispersions, dispersionTable, setOrderingFilter) to prepare the data for pseudotime analysis. The dimension is reduced using DDRTree and the cells are ordered.

Heatmaps

For each resolution, genes are tested for those that have differential expression as a result of pseudotime. They are then sorted by q value and the top 40 are selected. Clusters are annotated on the heatmap.

Monocle 3

For each resolution, the Seurat object is converted to a cell data set. Cells are then clusters and subset. learn_graph learns the principal graph. Cells are finally ordered.

Heatmaps

For each resolution, graph_tests for genes differentially expressed as a function of pseudotime. They are ordered by q-value and the top 40 selected. lusters are annotated on the heatmap.

g. Figure Overviews of the Workflow