Documentation

This is a single-cell RNA sequencing Shiny app built to better analyze and understand your scRNA-seq data. Upload your data and click on the different tabs to analyze your data clusters, individual genes, marker genes, or cell trajectories!

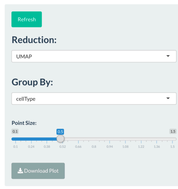
Upload Data

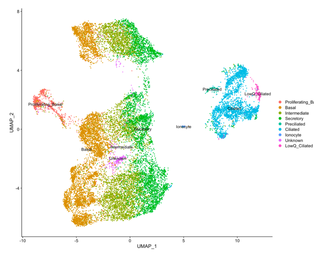
- Upload your Seurat data as an RDS object, a text-box will appear on the right describing the contents of your object
- Seurat Objects should have cells filtered, be scaled, and have QC/linear dimension reductions performed.

Data Clusters

Explore various cell types with different dimension reductions, coloring based on Seurat metadata factors:

Options:



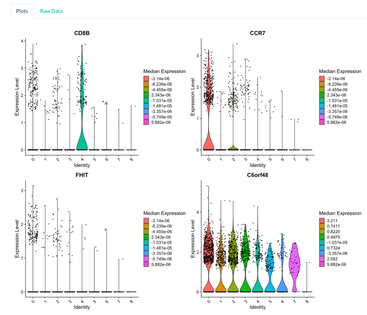


- Refresh: Update the plot on the right using the button after choosing different settings
- Dimension Reduction Type: will graph different linear dimension reductions performed on your Seurat object (PCA, TSNE, UMAP, etc.)
- Group By: will extract factors from your Seurat metadata to visualize (i.e. cell type, donor, and treatment performed on a cell)
- Download: Select your file type (pdf, png, jpeg, etc.) and plot image dimensions to download.

Gene Analyzer

Allow users to input interesting genes and visualize gene expression in various Seurat identities:

Options:



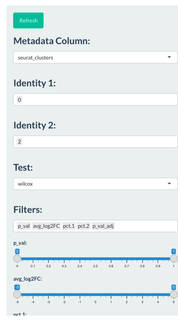
- Gene: This input box allows you to input which genes you would like to look at. Must have spaces or new lines in between gene names.
  - "HGNC" nomenclature genes and Gene symbols (i.e. not Ensemble ID etc.) are expected
- Metadata Column*: This selection determines which column of your metadata your identities are from
- Identities*: This selection determines which cluster identities you would like to compare
- Plot Type: This selection determines whether you would like to view your genes as a FeaturePlot, ViolinPlot, DotPlot, or HeatMap

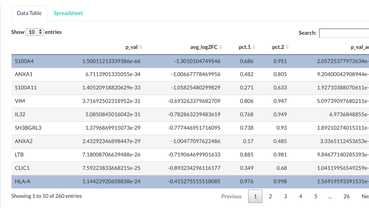
*for ViolinPlot, DotPlot, and HeatMap only

Marker Genes

Allow users to identify top marker genes for various cluster identities across other identity classes

Options:



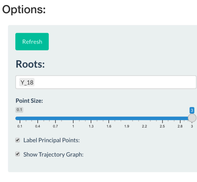


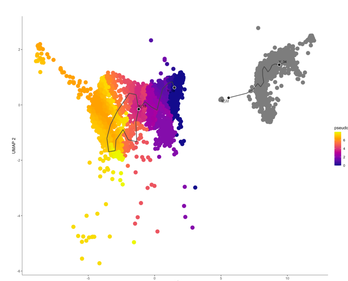
- Metadata Column: This selection determines which column of your metadata your identities are from (i.e. cell type, donor, or cytokine treatment performed on a cell)
- Identity 1: Choose the identity classes that you want to determine marker genes for (such as differentially expressed genes).
- Identity 2: Determines identity classes for comparison (ident1 and ident2 are interchangeable)
- Test: Denotes which comparison test to use. Available options are wilcox, bimod, roc, t, negbinom, poisson, LR, MAST, and DESeq2
  - "wilcox" : Identifies differentially expressed genes between two groups of cells using a Wilcoxon Rank Sum test (default)
  - "bimod" : Likelihood-ratio test for single cell gene expression, (McDavid et al., Bioinformatics, 2013)
  - "roc" : Identifies 'markers' of gene expression using ROC analysis. For each gene, evaluates (using AUC) a classifier built on that gene alone, to classify between two groups of cells. An AUC value of 1 means that expression values for this gene alone can perfectly classify the two groupings (i.e. Each of the cells in cells.1 exhibit a higher level than each of the cells in cells.2). An AUC value of 0 also means there is perfect classification, but in the other direction. A value of 0.5 implies that the gene has no predictive power to classify the two groups. Returns a 'predictive power' (abs(AUC-0.5) * 2) ranked matrix of putative differentially expressed genes.
  - "t" : Identify differentially expressed genes between two groups of cells using the Student's t-test.
  - "negbinom" : Identifies differentially expressed genes between two groups of cells using a negative binomial generalized linear model. Use only for UMI-based datasets
  - "poisson" : Identifies differentially expressed genes between two groups of cells using a poisson generalized linear model. Use only for UMI-based datasets
  - "LR" : Uses a logistic regression framework to determine differentially expressed genes. Constructs a logistic regression model predicting group membership based on each feature individually and compares this to a null model with a likelihood ratio test.
  - "MAST" : Identifies differentially expressed genes between two groups of cells using a hurdle model tailored to scRNA-seq data. Utilizes the MAST package to run the DE testing.
  - "DESeq2" : Identifies differentially expressed genes between two groups of cells based on a model using DESeq2 which uses a negative binomial distribution (Love et al, Genome Biology, 2014). This test does not support pre-filtering of genes based on average difference (or percent detection rate) between cell groups. However, genes may be pre-filtered based on their minimum detection rate (min.pct) across both cell groups.
- Filters: Uses slider inputs to select cutoffs for marker genes
  - Available filters include avg_log2FC, pct.1, pct.2, p_val, and p_val_adj
  - avg_log2FC: Average log2 fold change between your chosen identities
  - pct.1: The percentage of cells where the gene is detected in the first identity
  - pct.2: The percentage of cells where the gene is detected in the second identity
  - p_val_adj: Adjusted p-value, based on bonferroni correction using all genes in the dataset
- Marker Genes can be viewed as sortable data table or spreadsheet (to copy-paste gene names for the Gene Analyzer tab)
  - Data can also be downloaded as a .csv text file

Cell Trajectory

Allow users to perform single-cell trajectory analysis and visualize how cells transition between different cell states (using the Monocle 3 package)

Options:





- Roots: Selects roots or "beginnings" to your trajectory that are occupied by cells from early time points
- Label Principal Points: Show points on the principal graph that can serve as root nodes for the start of your trajectory
- Show Trajectory Graph: Displays the path cells travel on as they progress through pseudotime

For any further questions, feel free to contact Alex Sima at alex.sima@ucsf.edu