

Reference Mapping (Shiny App)

Introduction

This application is an R Shiny website for mapping single cell data to Reference data.

Selecting References

The index page of the app shows the list of references available. Pick a reference of choice from the dropdown menu

The screenshot shows the 'Reference Mapping' Shiny App interface. On the left sidebar, the 'Reference List' tab is active, and the 'Select a reference' dropdown menu is highlighted with a red box, showing 'LungMAP_MouseLungDevelopment_CellRef'. Below it, the 'Upload Query RDS File' section has a 'Browse...' button and a 'No file selected' message. The main panel displays 'Available Reference Datasets' with a search bar and a table of references. The table has columns for Reference, Description, and Species. The first row is highlighted in blue.

Reference	Description	Species
LungMAP_HumanLung_CellRef_Seeds.v1.1	LungMAP CellCards Reference human	human
LungMAP_MouseLungDevelopment_CellRef_Seeds.v1.1	LungMAP CellCards Reference mouse	mouse
GSE227136_ILD_all_celltypes_Seurat	Lung ILD Samples - all celltypes (GSE227136)	human
GSE227136_ILD_endothelial_Seurat	Lung ILD Samples - endothelial subset (GSE227136)	human
GSE227136_ILD_epithelial_Seurat	Lung ILD Samples - epithelial subset (GSE227136)	human
GSE227136_ILD_mesenchymal_Seurat	Lung ILD Samples - mesenchymal subset (GSE227136)	human
GSE227136_ILD_immune_Seurat	Lung ILD Samples - immune subset (GSE227136)	human

Upload query

Browse and upload the dataset to be annotated (query)

This screenshot is similar to the previous one, but the 'Upload Query RDS File' section in the sidebar is highlighted with a red box. It shows a 'Browse...' button and a file named 'FPR_I73T_Mesenchyme.RDS' with an 'Upload complete' button next to it. The main panel remains the same, showing the 'Available Reference Datasets' table.

Selecting input arguments to run mapping

In the side panel, click the preprocessing tab. It will populate all input variables for running the Seurat's reference mapping. Choose input arguments.

- *Normalization method*: Name of normalization method used- LogNormalize or SCT

- *Select a reference assay:* Name of the Assay to use from reference
- *Select a query assay:* Name of the Assay to use from query.
- *Recompute residuals:* If using SCT as a normalization method, compute query Pearson residuals using the reference SCT model parameters
- *Dimension to use from reduction:* number of principle components (dimensions) to use
- *Number of trees to use in nearest neighbor search:* Number of trees to use in nearest neighbor method. More trees gives higher precision when using annoy approximate nearest neighbor search
- *Reference metadata to transfer:* columns present in the metadata that you would like to use for labelling the query dataset. You can choose multiple options
- *Reference reduction:* Name of dimensional reduction to use from the reference if running the pcaproject workflow

Reference Mapping

Reference List

Select a reference

LungMAP_MouseLungDevelopment_CellRef

Upload Query RDS File

Browse... FPR_IT3T_Mesenchyme.RDS

Upload complete

Mapping

>> Preprocessing

>> Feature Plots

Enter inputs for finding anchors and Map query

Normalization method

LogNormalize

Select a reference assay

RNA

Select a query assay

RNA

Recompute residuals

FALSE

Dimension to use from reduction

10 50 100

Number of trees to use in nearest neighbor search

5 20 50

Reference metadata to transfer

lineage_level1 celltype_level1 celltype_level2

Reference reduction

pca

Map cells to reference

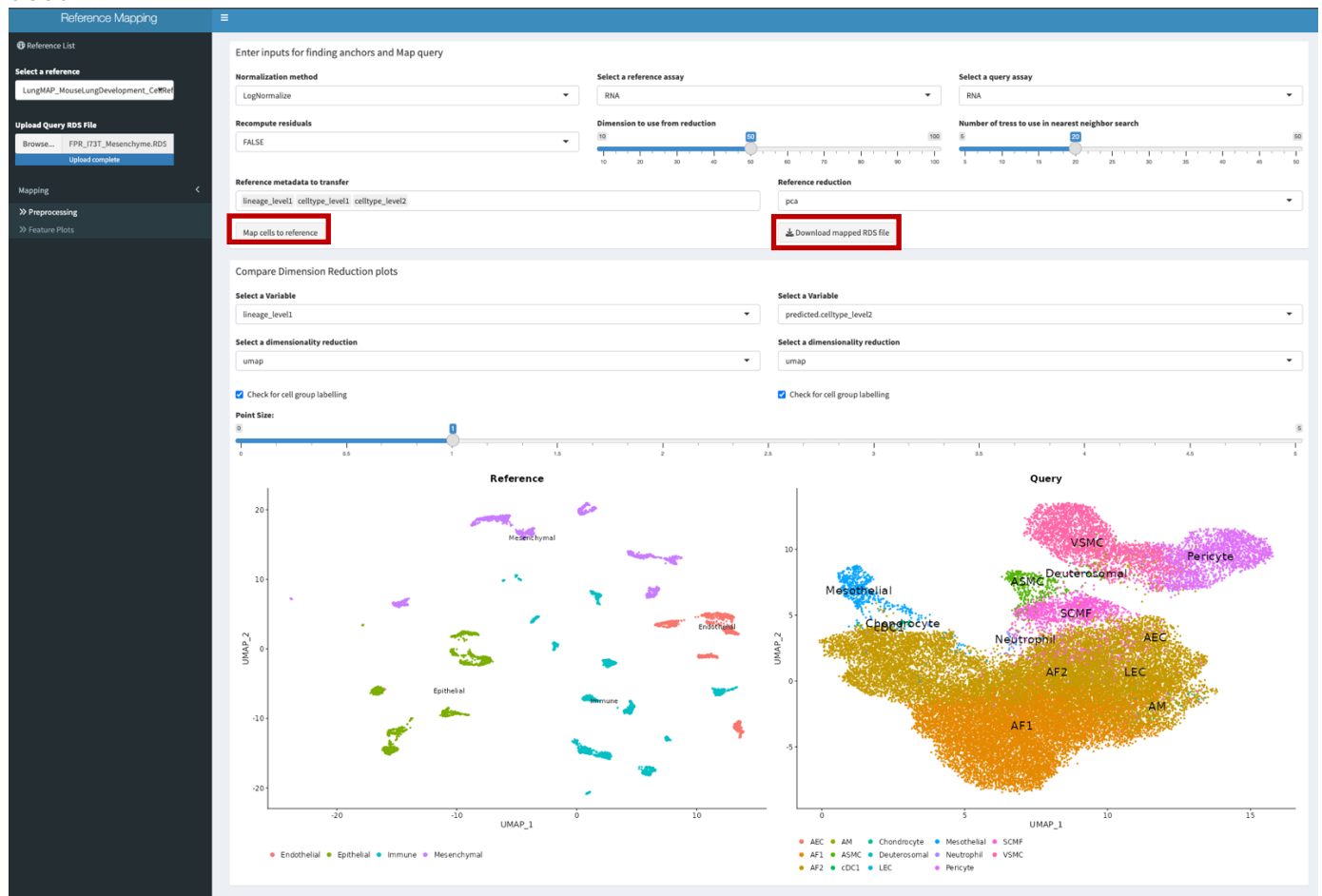
Download mapped RDS file

Mapping cells to reference

Once the input selections are made, including what metadata from the reference needs to be transferred from the query, click on the *Map cells to reference* button (this may take a while). Once the reference annotations are mapped, the Seurat object with the mapped annotations can be downloaded as an RDS file using the *Download mapped RDS file* button.

You can also visualize the results by looking at the Dimension Reduction plots below the control panel. On the left is the reference plots and on the right is the query plots. You can use the dropdown menu to choose which reduction you want to see (pca/tsne/umap) and then also select the metadata column that you want the data to be grouped by. For example, in the screenshot below, Umap is the dimension reduction of choice and while the reference dataset displays the lineage at level 1, the query displays the predicted celltype at level 2.

Please keep in mind that lineage_level1 and celltype_level2 are just how columns are organized in the reference metadata and they may vary depending on the reference being used.



Feature Plots

The Feature plots menu in the sidebar has 2 options

- Gene : You can look at gene expression levels in the dataset by selecting a gene from the dropdown menu
- Prediction scores : Though the program predicts cell type based on reference, you can look at how well it does so but looking at the prediction scores

