

Automated MAIT cell annotation in Lett dataset using SingleR

scRNA-seq workflow series part3

Ivan Osinnii

2024-04-18

Table of contents

1 Load data and dependencies	2
2 Analysis with SingleR and MonacoImmuneData reference	3
2.1 Validate annotation using quality control plots	3
2.2 Make projection of all MonacoImmuneData cell types on Lett dataset UMAP plot	5
2.3 Next we want to do some adjusted visualizations to focus only on MAIT cells	6
2.3.1 Use ifelse function to create a new factor vector with only MAIT label and “others” and plot	6
3 Analysis of Lett dataset using SingleR and Garner et al. dataset for reference	8
3.0.1 Extract count matrix from seurat object for ref data	9
3.0.2 Let’s construct reference object from Garner dataset	9
3.0.3 Make second prediction model based on Garner reference	10

1 Load data and dependencies

```
library(SingleCellExperiment)
library(SummarizedExperiment)
library(Seurat)
library(SingleR)
library(pheatmap)
library(ggplot2)
library(RColorBrewer)
library(tidyverse)
library(scater)
library(uwot)
library(celldex)

getwd()

[1] "/Users/osinnii/Documents/Github/scRNA-seq-workflow/scRNA-seq-workflow/R"

lett.sce <- readRDS("./input/Lett.sce.liver.gene.symbol_analyzed.rds")
lett.sce

class: SingleCellExperiment
dim: 36169 6724
metadata(3): Samples Samples Samples
assays(5): counts genefull spliced unsPLICED logcounts
rownames(36169): PLCXD1 LINC00685 ...
rowData names(12): GENEID SYMBOL ... max_prop_ambient topHVG
colnames(6724): BSSE_QGF_176995-AAACCCAAGACTCCGC
  BSSE_QGF_176995-AAACCCAAGATACTAGT ...
  BSSE_QGF_176995-TTTGTTGTCGTTAGTG
colData names(23): Barcode SampleName ... MonImmCell label
reducedDimNames(7): PCA TSNE ... UMAP UMAP.2
mainExpName: RNA
altExpNames(0):
```

2 Analysis with SingleR and MonacoImmuneData reference

Extract count matrix from sce object for query data and load reference

```
query <- assay(lett.sce)

ref <- celldex::MonacoImmuneData()
```

Make a prediction model using SingleR function

```
pred <- SingleR(test = query, ref = ref, labels = ref$label.fine)

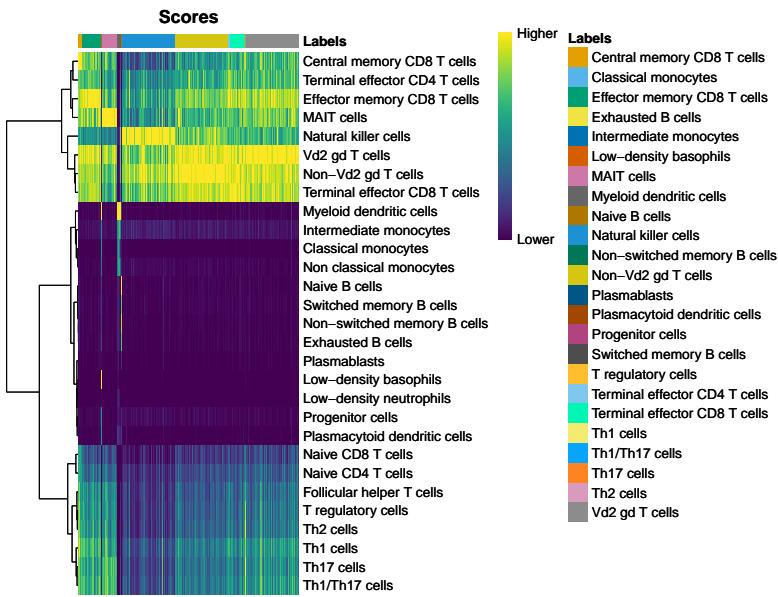
pred.Tcell <- SingleR(test = query, ref = ref, labels = ref$label.main)
```

2.1 Validate annotation using quality control plots

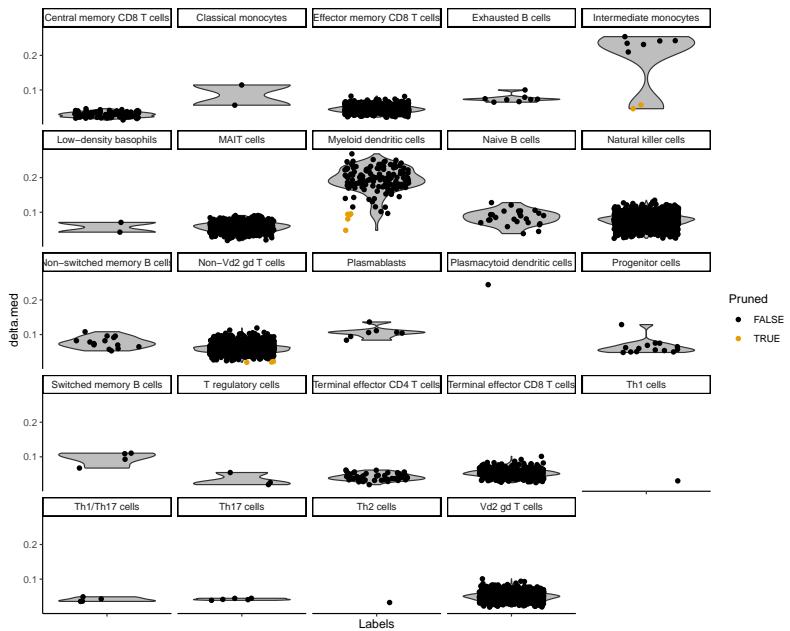
```
pred$scores[7:9,7:9]
```

	Low-density basophils	Low-density neutrophils	MAIT cells
[1,]	0.1390851	0.1372944	0.2354845
[2,]	0.1353154	0.1334808	0.2529216
[3,]	0.1443342	0.1520688	0.3169705

```
plotScoreHeatmap(pred)
```

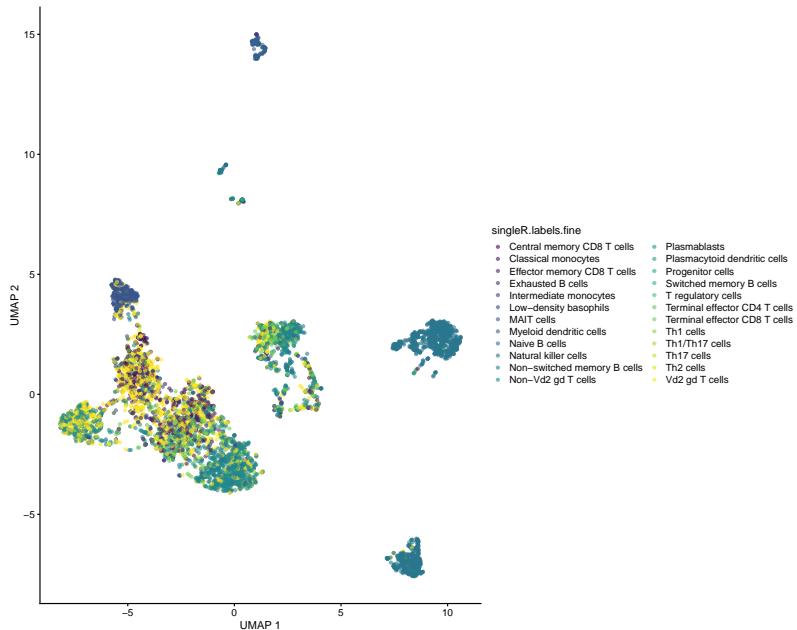


plotDeltaDistribution(pred)

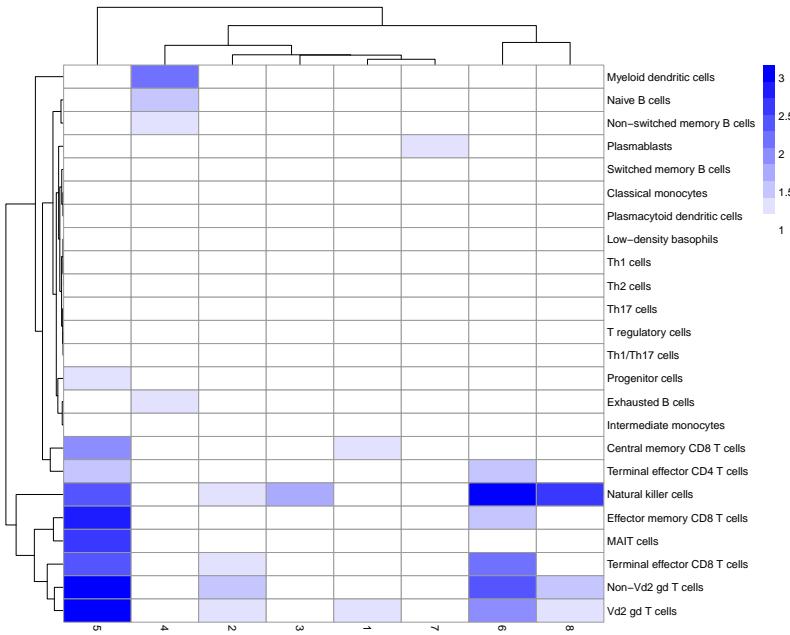


2.2 Make projection of all MonacolimmuneData cell types on Lett dataset UMAP plot

```
lett.sce$singleR.labels.fine <- pred$labels[match(rownames(lett.sce@colData),  
rownames(pred))]  
  
plotUMAP(lett.sce, colour_by = 'singleR.labels.fine',  
point_size = 1.0)
```



```
tab <- table(Assigned=pred$labels, Clusters=lett.sce$label)  
  
pheatmap(log10(tab+10), color = colorRampPalette(c('white','blue'))(10))
```



2.3 Next we want to do some adjusted visualizations to focus only on MAIT cells

2.3.1 Use ifelse function to create a new factor vector with only MAIT label and “others” and plot

```
cell_type_of_interest <- "MAIT cells"

MAIT_cells <- ifelse(lett.sce@colData@listData[["singleR.labels.fine"]] == "MAIT cells", "MAIT cells", "All other")

colData(lett.sce)$MAIT_cells <- factor(MAIT_cells)

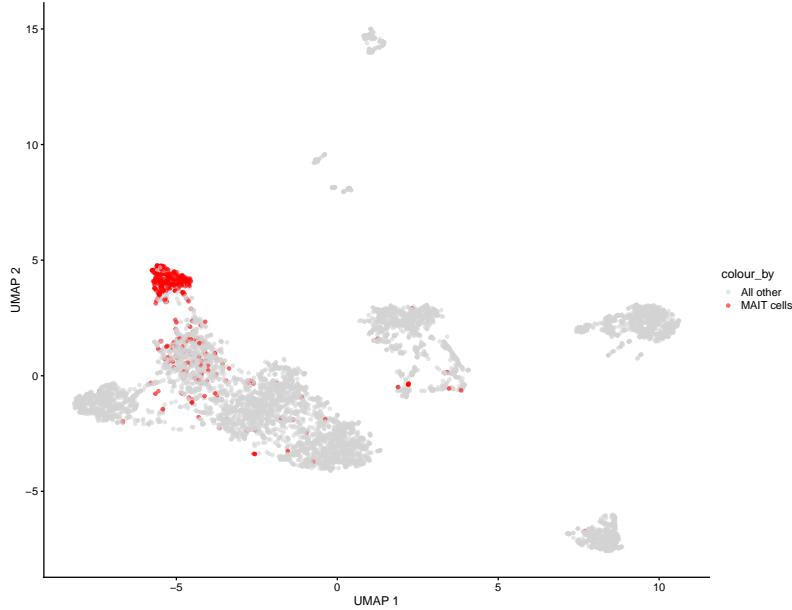
colors <- c("MAIT cells" = "red", "All other" = "lightgrey")

p <- plotReducedDim(lett.sce, "UMAP", colour_by = "MAIT_cells", point_size = 1) + scale_col
```

Scale for colour is already present.

Adding another scale for colour, which will replace the existing scale.

```
print(p)
```



Then we want to see MAIT cell counts compared to total T cells. This is achieved by next code section.

```
# Let's count MAIT cells and Other cells
MAIT_cell_values <- colData(lett.sce)$MAIT_cells
value_counts <- table(MAIT_cell_values)

# now let's annotate all T cells

lett.sce$singleR.labels.main <- pred.Tcell$labels[match(rownames(lett.sce@colData),
rownames(pred.Tcell))]

# Filtering out only "T cells"
t_cell_values <- lett.sce$singleR.labels.main
is_t_cell <- t_cell_values == "T cells"
t_cell_counts <- sum(is_t_cell) # Count the number of "T cells"
percent1 <- round((value_counts[2] / t_cell_counts * 100),1)

message("Number of MAIT cells equals ", value_counts[2])
```

```
Number of MAIT cells equals 459
```

```
message("Number of all T cells equals ", t_cell_counts)
```

```
Number of all T cells equals 4375
```

```
message("Percentage of MAIT cells out of all T cells found with this method of annotation eq
```

```
Percentage of MAIT cells out of all T cells found with this method of annotation equals 10.5
```

3 Analysis of Lett dataset using SingleR and Garner et al. dataset for reference

Let's load the data

```
ref2 = readRDS("output/Garner.seurat.2exp.liver.mait_analyzed.rds")
```

This command would be needed to convert it from Seurat to SCE, but we can avoid it (see later)

```
#sceasy::convertFormat(Gar.sce,
# from = "seurat", to = "sce",
# outFile = "Gar_sce.rds")

# Gar_sce is garner.liver.withTrm converted into sce

#Gar.sce <- readRDS("Gar_sce.rds")
# Gar.sce
# colLabels(Gar.sce) <- Gar.sce$RNA_snn_res.0.1
# plotUMAP(Gar.sce, colour_by = "label")
```

3.0.1 Extract count matrix from seurat object for ref data

```
ref2.counts <- GetAssayData(ref2, layer = "data")  
  
common_genes <- intersect(rownames(lett.sce), rownames(ref2))  
length(common_genes)
```

```
[1] 15243
```

```
# Subset lett.sce to include only common genes  
lett.sce <- lett.sce[common_genes, ]  
  
ref2 <- ref2[common_genes, ]  
ref2
```

```
An object of class Seurat  
15243 features across 17069 samples within 1 assay  
Active assay: RNA (15243 features, 1717 variable features)  
3 layers present: counts, data, scale.data  
2 dimensional reductions calculated: pca, umap
```

```
query <- assay(lett.sce)
```

3.0.2 Let's construct reference object from Garner dataset

```
cell_metadata <- ref2$cell_type  
  
metadata_df <- DataFrame(cell_type = cell_metadata)  
  
ref_se_object <- SummarizedExperiment(assays = ref2.counts,  
                                         colData = cell_metadata)  
  
ref_se_object$cell_type <- ref_se_object$X  
  
ref_se_object@assays@data@listData[["logcounts"]] <- ref_se_object@assays@data@listData[[1]]
```

```
ref_se_object
```

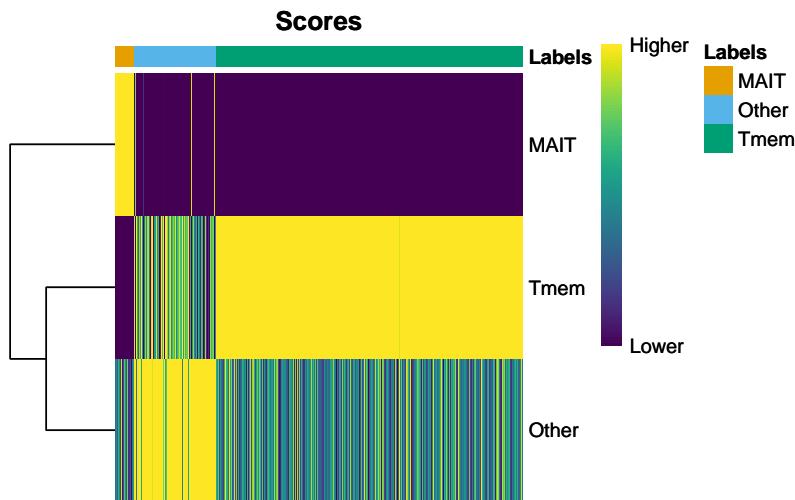
```
class: SummarizedExperiment
dim: 19289 17069
metadata(0):
assays(2): '' logcounts
rownames(19289): AL627309.1 FAM87B ... AC233755.1 AC240274.1
rowData names(0):
colnames(17069): 1_AAAGATGAGTTGAGAT 1_AAATGCCGTGCTCTTC ...
8_TTTGTCAGTGCATCTA 8_TTTGTCATCTTAGAGC
colData names(2): X cell_type
```

3.0.3 Make second prediction model based on Garner reference

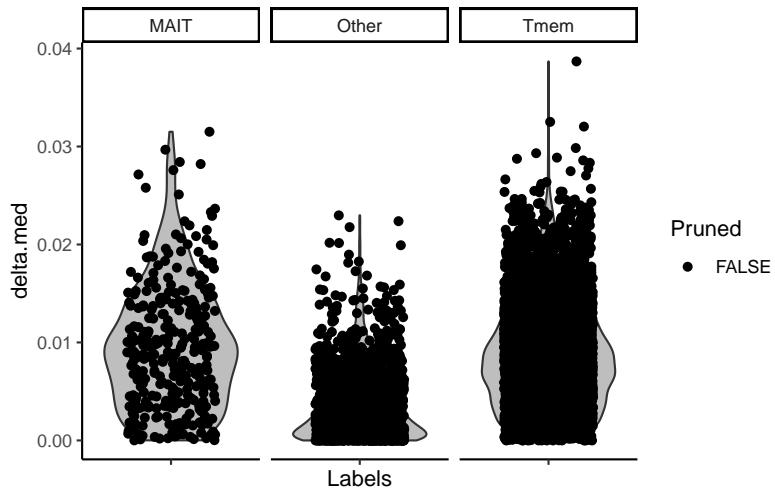
```
pred2 <- SingleR(test = query, ref = ref_se_object, labels = ref_se_object$cell_type)
```

4 Annotation “QC” plots

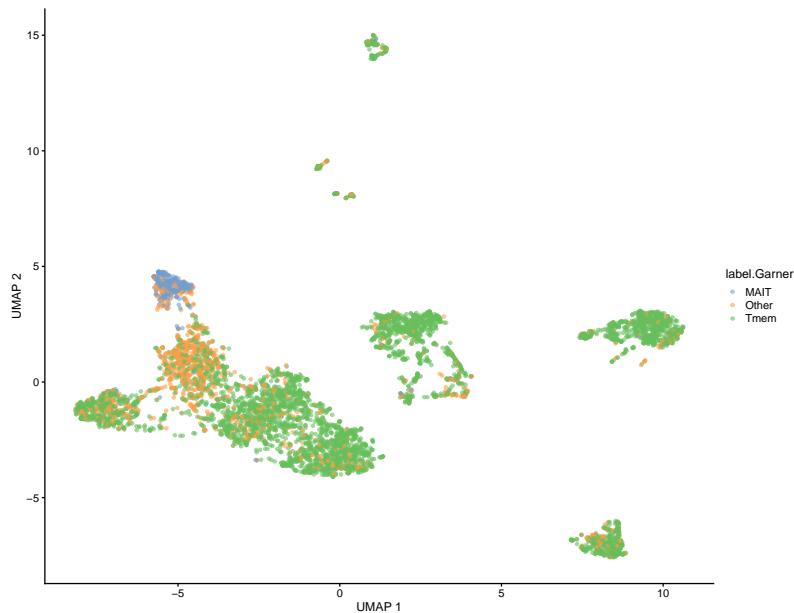
```
plotScoreHeatmap(pred2)
```



```
plotDeltaDistribution(pred2)
```



```
lett.sce$label.Garner <- pred2$labels[match(rownames(lett.sce@colData), rownames(pred2))]  
plotUMAP(lett.sce, colour_by = 'label.Garner', point_size = 1.0)
```



4.0.0.1 Use ifelse function to create a new factor vector with only MAIT label and “others”

```
#cell_type_of_interest <- "MAIT cells"

MAIT_cells <- ifelse(lett.sce@colData@listData[["label.Garner"]] == "MAIT", "MAIT", "All oth

colData(lett.sce)$MAIT_cells_Garner <- factor(MAIT_cells)

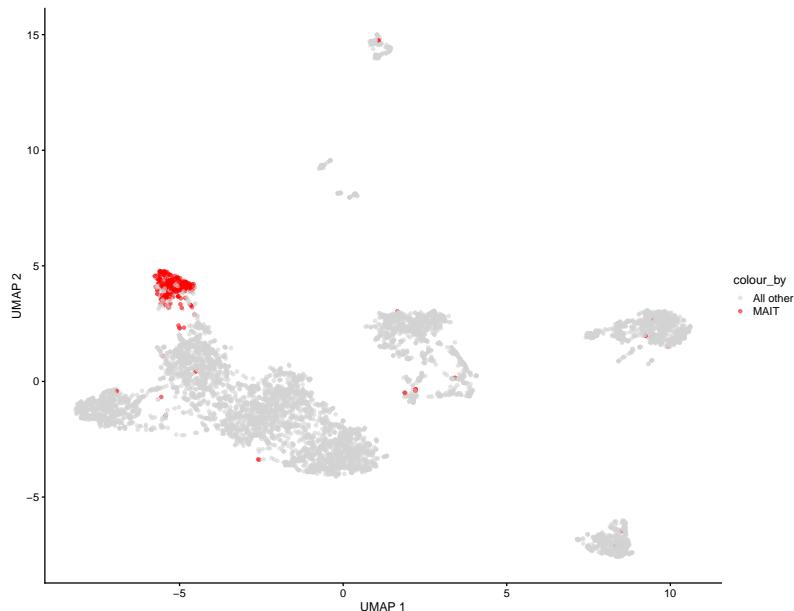
colors <- c("MAIT" = "red", "All other" = "lightgrey")

p2 <- plotReducedDim(lett.sce, "UMAP", colour_by = "MAIT_cells_Garner", point_size = 1) + sc
```

Scale for colour is already present.

Adding another scale for colour, which will replace the existing scale.

```
print(p2)
```



Calculating MAIT cell number and ratio to all T cells

```
MAIT_cell_values2 <- colData(lett.sce)$MAIT_cells_Garner  
value_counts2 <- table(MAIT_cell_values2)  
percent2 <- round((value_counts2[2] / t_cell_counts * 100),1)  
message("Number of MAIT cells equals ", value_counts2[2])
```

Number of MAIT cells equals 324

```
message("Number of all T cells equals ", t_cell_counts)
```

Number of all T cells equals 4375

```
message("Percentage of MAIT cells out of all T cells found with this method of annotation equals ", percent2)
```

Percentage of MAIT cells out of all T cells found with this method of annotation equals 7.4