Introduction to scRNA-seq integration

Compiled: October 22, 2023

scRNA-seq Integration with GABA1 and GABA2 data sets

```
library(Seurat)
library(patchwork)
```

Create Seurat objects from the filtered cell ranger data

```
experiment_name <- "C. elegans scRNA-seq"
dataset_loc <- "../ackley_data"
ids <- c("GABA1", "GABA2")

d10x.data <- lapply(ids, function(i) {
    data <- Read10X(data.dir = file.path(dataset_loc, i, "outs", "filtered_feature_bc_matrix"))
})
names(d10x.data) <- ids

sobject.list <- lapply(ids, function(i) {
    sobject <- CreateSeuratObject(counts = d10x.data[[i]])
    # remember which id the data belongs to (useful after integration)
    sobject[["group"]] <- rep(i, times = length(sobject@meta.data$nCount_RNA))
    # Does the 'MT' prefix signify mitochondrial DNA in C. elegans and does it need to be
    # filtered out?
    sobject[["percent.mt"]] <- PercentageFeatureSet(sobject, pattern = "^MT")
    sobject
})
names(sobject.list) <- ids
```

Normalization

```
# normalize and identify variable features for each dataset independently
sobject.list <- lapply(X = sobject.list, FUN = function(x) {
    x <- NormalizeData(x)
    x <- FindVariableFeatures(x, selection.method = "vst", nfeatures = 2000)
})

# select features that are repeatedly variable across datasets for integration
features <- SelectIntegrationFeatures(object.list = sobject.list)
```

Perform integration

```
grouped.anchors <- FindIntegrationAnchors(object.list = sobject.list, anchor.features = features)
```

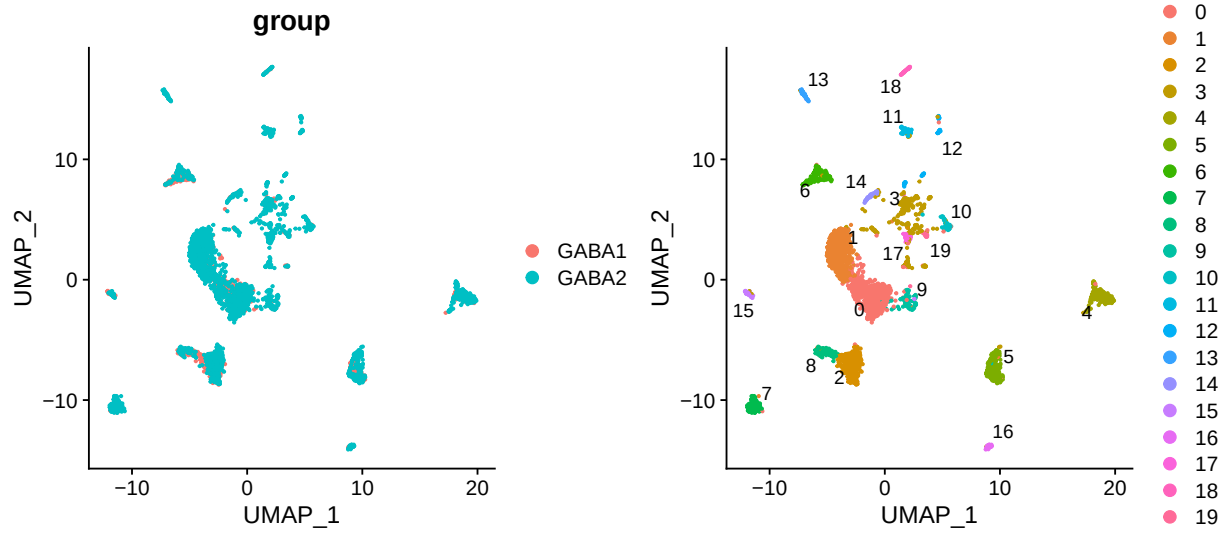
```
# this command creates an 'integrated' data assay
grouped.combined <- IntegrateData(anchorset = grouped.anchors)
```

Perform an integrated analysis

Now we can run a single integrated analysis on all cells!

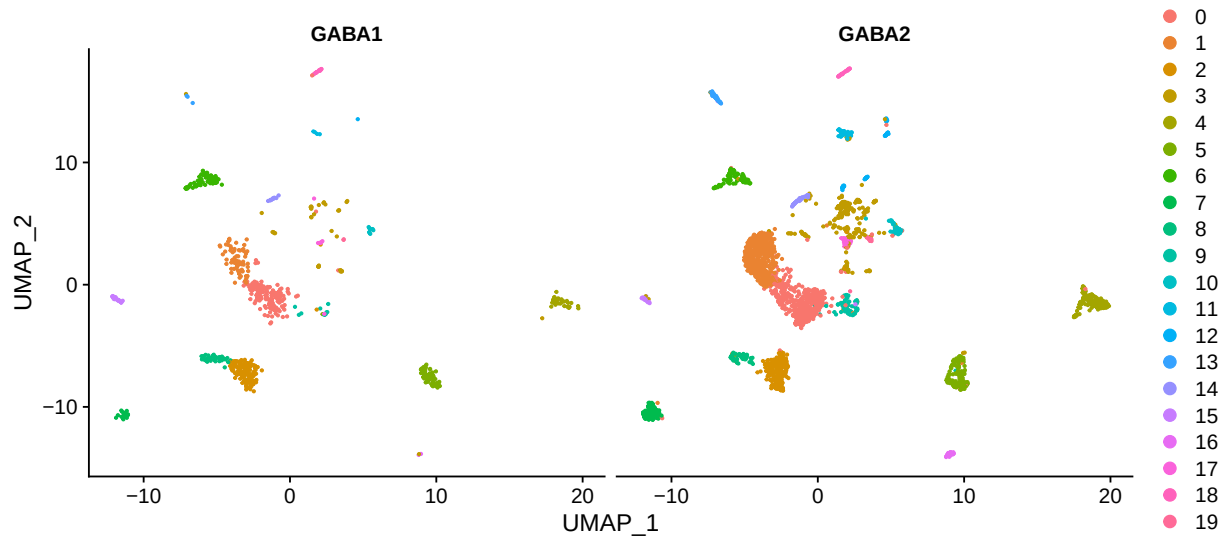
```
# specify that we will perform downstream analysis on the corrected data note that the
# original unmodified data still resides in the 'RNA' assay
DefaultAssay(grouped.combined) <- "integrated"

# Run the standard workflow for visualization and clustering
grouped.combined <- ScaleData(grouped.combined, verbose = FALSE)
grouped.combined <- RunPCA(grouped.combined, npcs = 30, verbose = FALSE)
grouped.combined <- RunUMAP(grouped.combined, reduction = "pca", dims = 1:30)
grouped.combined <- FindNeighbors(grouped.combined, reduction = "pca", dims = 1:30)
grouped.combined <- FindClusters(grouped.combined, resolution = 0.5)
```

```
# Visualization
p1 <- DimPlot(grouped.combined, reduction = "umap", group.by = "group")
p2 <- DimPlot(grouped.combined, reduction = "umap", label = TRUE, repel = TRUE)
p1 + p2
```



To visualize the two conditions side-by-side, we can use the `split.by` argument to show each condition colored by cluster.

```
DimPlot(grouped.combined, reduction = "umap", split.by = "group")
```



Identify conserved cell type markers

```
# For performing differential expression after integration, we switch back to the original
# data
DefaultAssay(grouped.combined) <- "RNA"
nk.markers <- FindConservedMarkers(grouped.combined, ident.1 = 6, grouping.var = "group", verbose = FALS
head(nk.markers)
```

```
##            GABA1_p_val GABA1_avg_log2FC GABA1_pct.1 GABA1_pct.2 GABA1_p_val_adj
## flp-10   8.658042e-151         5.921891       1.000       0.015   4.062700e-146
## vab-23    1.775101e-86         2.742712       0.580       0.004    8.329485e-82
## nlp-40   4.223174e-101         5.859277       1.000       0.107    1.981682e-96
## D1007.19 5.318342e-137         4.355848       0.943       0.018   2.495579e-132
## T05A8.3   1.289278e-76         6.213582       1.000       0.223    6.049808e-72
## srv-7     4.384392e-66         2.211071       0.432       0.001    2.057332e-61
##            GABA2_p_val GABA2_avg_log2FC GABA2_pct.1 GABA2_pct.2 GABA2_p_val_adj
## flp-10    0.000000e+00         4.651373       0.877       0.008    0.000000e+00
## vab-23   2.797923e-198         2.533769       0.438       0.003   1.312897e-193
## nlp-40   6.358671e-198         4.830521       1.000       0.057   2.983743e-193
## D1007.19 1.812119e-150         3.346140       0.699       0.032   8.503187e-146
## T05A8.3  5.047622e-148         6.947285       1.000       0.093   2.368546e-143
## srv-7    1.806037e-118         1.888279       0.260       0.002   8.474649e-114
##               max_pval minimump_p_val
## flp-10   8.658042e-151   0.000000e+00
## vab-23    1.775101e-86  5.595845e-198
## nlp-40   4.223174e-101  1.271734e-197
## D1007.19 5.318342e-137  3.624238e-150
## T05A8.3   1.289278e-76  1.009524e-147
## srv-7     4.384392e-66  3.612074e-118
```