

# Cluster Analysis of SCI Immune Cells

James Choi

Last run: 2021-02-10

```
knitr::opts_chunk$set(warning=FALSE, message=FALSE, fig.align='center',
tidy.opts=list(width.cutoff=60), tidy=TRUE)
```

## Cluster Analysis of SCI Myeloid cells

The goal of this analysis is to identify subsets of immune cells after SCI and to determine which are shared or unique to the injury responses between neonatal and adult mouse models.

```
if (!grepl("scripts", getwd())) {
  setwd("./scripts/")
}
require("Seurat")
require("ggplot2")
require("dplyr")
require("cowplot")

results_out <- ".../results/ClusterAnalysisAllMyeloid/"
dir.create(results_out)

# sci <- readRDS(file = '../data/myeloid_combined.rds')
```

First, we cluster the cells. We use the graph-based louvain clustering approach that is implemented in Seurat. We use default settings and elect to merge non-distinct or consider finer clustering after initial identification of cell-types and subsets.

```
DefaultAssay(sci) <- "integrated"
sci <- FindClusters(sci, resolution = 0.8, verbose = FALSE)
p1 <- DimPlot(sci, label = TRUE, label.size = 5) + theme_bw() +
  NoLegend()
p2 <- DimPlot(sci, group.by = "study", shuffle = TRUE) + theme_bw()
p3 <- DimPlot(sci, split.by = "time") + theme_bw() + NoLegend()
p4 <- plot_grid(plot_grid(p1, p2, ncol = 2, rel_widths = c(0.8,
  1)), p3, ncol = 1, rel_heights = c(2, 1))
ggsave(filename = paste0(results_out, "allMyeloid_cluster-summary_umap.tiff"),
  plot = p4, height = 6, width = 9, device = "tiff")
p4
```

Inspecting the UMAPs, we see that there are some age-specific immune cells as well as subsets that are enriched in neonatal vs. adult. For example, cluster 13 and cluster 17 are specific to adult and neonatal, respectively, while many more cells in cluster 12 come from the neonatal dataset. Cluster 10 appears to be

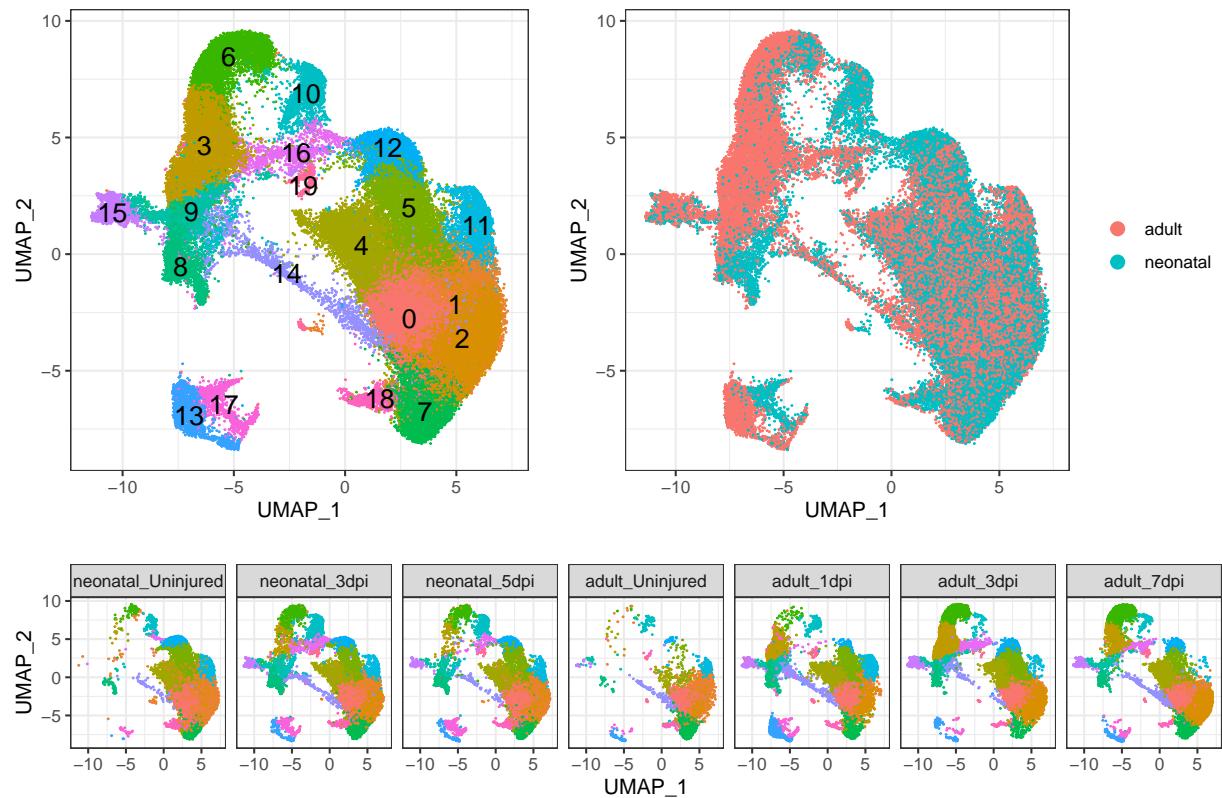


Figure 1: UMAP of SCI immune cells. Top left: cells colored by cluster. Top right: cells colored by study of origin. Bottom: cells colored by cluster and split by injury time-point across both studies.

comprised almost entirely from neonatal dataset. To better characterize these clusters, we next identify the globally differentially expressed genes per cluster and examine the top few.

```
markers <- FindAllMarkers(object = sci, assay = "RNA", slot = "data",
  test.use = "wilcox", logfc.threshold = log(2), # only.pos = TRUE)
write.table(x = markers, file = paste0(results_out, "defaultClusterMarkers_minFClog2.txt"),
  sep = "\t", quote = FALSE, row.names = FALSE)

markers <- read.table(file = paste0(results_out, "defaultClusterMarkers_minFClog2.txt"),
  header = TRUE)
top_markers <- markers %>% group_by(cluster) %>% filter(avg_logFC >
  0) %>% top_n(n = 3, wt = -p_val_adj) %>% top_n(n = 3, wt = avg_logFC)
knitr::kable(x = top_markers, caption = "Top 3 differentially expressed genes per cluster")
```

Table 1: Top 3 differentially expressed genes per cluster

p_val	avg_logFC	pct.1	pct.2	p_val_adj	cluster	gene
0	0.8511427	0.970	0.674	0	0	Cd83
0	0.7938463	0.796	0.463	0	0	Tnf
0	0.7762550	0.900	0.716	0	0	Ccl4
0	0.9084122	0.929	0.646	0	1	P2ry12
0	0.8168659	0.753	0.420	0	1	Crybb1
0	0.7630616	0.859	0.490	0	1	Siglech
0	0.7648233	1.000	0.995	0	2	Cst3
0	0.7464414	1.000	0.818	0	2	Cd81
0	0.7315113	0.981	0.750	0	2	Olfml3
0	2.5141011	0.912	0.553	0	3	Fabp5
0	2.1704705	0.814	0.208	0	3	Gpnmb
0	2.1187115	0.999	0.510	0	3	Lgals3
0	1.1409207	0.857	0.447	0	4	Lpl
0	0.9518011	0.645	0.300	0	4	Igf1
0	0.7033592	0.997	0.900	0	4	Cd63
0	0.9897317	0.855	0.595	0	5	Ccl12
0	0.9572212	0.855	0.418	0	5	Stmn1
0	0.7745166	0.342	0.147	0	5	Ube2c
0	1.7585593	0.984	0.883	0	6	Lyz2
0	1.5621337	0.701	0.235	0	6	Gpnmb
0	1.4675586	0.943	0.252	0	6	Ms4a7
0	1.1474710	0.966	0.833	0	7	Jun
0	1.1090365	0.897	0.646	0	7	Klf2
0	1.1056937	0.949	0.735	0	7	Rhob
0	2.6610933	0.941	0.114	0	8	Plac8
0	1.7460904	0.997	0.551	0	8	Ifitm3
0	1.6765138	0.959	0.251	0	8	S100a4
0	2.1284585	0.708	0.152	0	9	Arg1
0	1.9366658	0.693	0.171	0	9	AA467197
0	1.9123164	0.964	0.513	0	9	Crip1
0	2.5460851	0.969	0.178	0	10	Mrc1
0	2.3098738	0.994	0.311	0	10	Pf4
0	1.9177586	0.956	0.201	0	10	F13a1
0	0.6999600	0.771	0.601	0	11	Ppp1r14b
0	1.4675289	0.772	0.610	0	11	Ccl12
0	1.2691806	0.364	0.208	0	11	Nme2

p_val	avg_logFC	pct.1	pct.2	p_val_adj	cluster	gene
0	1.6963506	0.998	0.436	0	12	Stmn1
0	1.6350657	0.885	0.095	0	12	Top2a
0	1.5426018	0.897	0.093	0	12	Pclaf
0	4.8031839	0.859	0.079	0	13	Retnlg
0	4.7218425	0.990	0.305	0	13	S100a9
0	4.4000981	0.995	0.393	0	13	S100a8
0	2.1313031	0.880	0.178	0	14	Isgr15
0	1.6299831	0.493	0.031	0	14	Rsd2
0	1.6015057	0.822	0.128	0	14	Irf7
0	3.8984208	0.966	0.173	0	15	H2-Ab1
0	3.8961433	0.969	0.083	0	15	H2-Aa
0	3.8697985	0.994	0.409	0	15	Cd74
0	1.2286550	0.944	0.445	0	16	Stmn1
0	1.1256775	0.991	0.582	0	16	Lgals1
0	1.1029276	0.686	0.108	0	16	Pclaf
0	3.5175678	0.527	0.062	0	17	Camp
0	3.3605373	0.562	0.069	0	17	Ngp
0	3.2813487	0.659	0.063	0	17	Stfa1
0	0.8058458	0.997	0.992	0	18	Malat1
0	0.7771464	0.959	0.936	0	18	mt-Nd2
0	0.7505881	0.914	0.806	0	18	mt-Nd3
0	3.3336972	0.346	0.012	0	19	Igkc
0	2.3749739	0.275	0.004	0	19	Elane
0	2.2721702	0.456	0.036	0	19	Prtn3

Because this test is a global comparison, it would benefit to generate violin plots to visualize distribution of gene expression.

```
top_gene <- markers %>% group_by(cluster) %>% top_n(n = 2, wt = -p_val_adj) %>%
  top_n(n = 2, wt = avg_logFC)
top_gene <- unique(top_gene$gene)

sci$seurat_clusters <- sci$integrated_snn_res.0.8
Idents(sci) <- "seurat_clusters"
DefaultAssay(sci) <- "RNA"
expr_dat <- FetchData(object = sci, vars = c("seurat_clusters",
  top_gene), slot = "data")
top_gene_vln <- expr_dat %>% reshape2::melt(id.vars = "seurat_clusters") %>%
  ggplot(mapping = aes(x = seurat_clusters, y = value)) + geom_violin(mapping = aes(fill = seurat_clusters,
  scale = "width") + facet_grid(variable ~ ., scales = "free_y") +
  xlab(label = "Cluster") + ylab(label = "log-normalized expression") +
  theme_bw() + theme(legend.position = "none")
ggsave(filename = paste0(results_out, "allMyeloid_default-cluster_top2genes_vln.tiff"),
  plot = top_gene_vln, device = "tiff", height = 20, width = 6)
top_gene_vln
```

Based on the previous marker gene violin plots and UMAP, clusters 3, 6, 8, 9, 13, 15, 16, and 17 correspond to peripherally-derived myeloid cells. Cluster 13 and 17 are subsets of neutrophils, but what distinguishes them needs to be further investigated. Cluster 8 is Ccr2+/Ly6c2+ monocyte. Cluster 9 is macrophages early in differentiation from monocytes and express Arg1 and Fn1. Accordingly, cluster 3 is likely a more “mature” macrophage in transition to becoming a foamy macrophage as seen by expression of Fabp4, Fabp5,

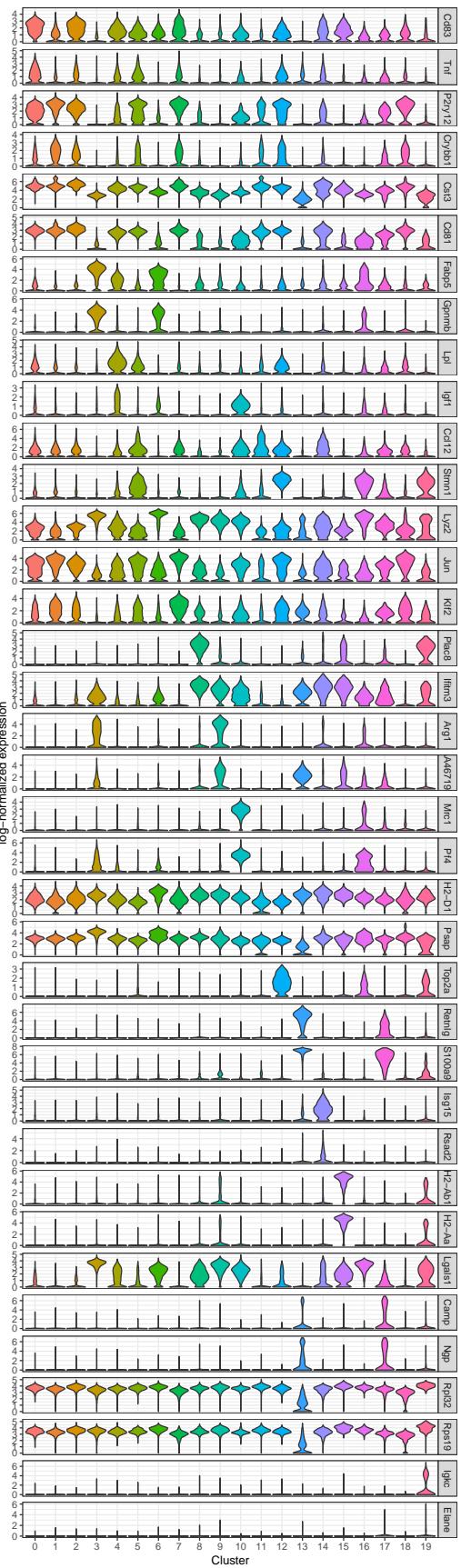


Figure 2: Violin plot of top 2 marker genes per cluster.  
5

and Plin2. Clusters 9 and 3 together likely comprise what we' previously described as macrophage-A. Not surprisingly, cluster 6 likely corresponds to macrophage-B. It is enriched in Gpnmb, Apoe, Ms4a7, and Lyz2. Cluster 15 highly expresses Cd74 and other MHC-class related genes. These are likely what we previously described as Dendritic cells. Cluster 16 expresses cell division genes.

Clusters 0, 1, 2, 4, 5, 7, 11, 12, and 18 are likely different subsets of microglia. Upregulated of Tnf and Cd83 by cluster 0 suggests an inflammatory and/or “activated” signature. P2ry12 is enriched in cluster 1, indicating a homeostatic subset. Cluster 2 is very similar to cluster 0 with few DE genes - possibly separated due to library size differences. Cluster 4 shows upregulation of Igf1, Lpl, and Spp1, which we previously identified to be a potentially [regeneration permissive subset of microglia](#) peaking later at 7dpi in adult SCI. Clusters 5 and 12 highly express genes related to proliferation such as Mki67 and Cenpa. Cluster 11 is enriched for genes such as Nme2, Ran, and ribosomal genes and is low in expression of Cx3cr1. Malat1 and mitochondrial genes are upregulated in cluster 18, which means it is probably a low-quality cluster that should be removed. Cluster 7 expresses immediate-early genes such as Jun, Klf2, and Fos. This signature is similar to the ex-DAMs found in a [report describing dissociation-induced effects on glial cells](#).

Cluster 19 expresses Igkc, which is a B cell marker. Cluster 14 is an immune subset enriched for interferon-pathway genes, which has been described in [other studies](#) but their role in SCI is unclear . Cluster 10 are likely border-associated macrophages as described by [Van Hove et al 2018](#).

After identifying marker genes per cluster, we can also identify how much of each cluster comprise each of the samples. We can further cluster the samples by similarity of cell-type composition to understanding similarities and differences in total immune cell response to SCI in adult vs neonatal.

```
clust_prop <- proportions(x = table(sci$integrated_snn_res.0.8,
  sci$sample_id), margin = 2) * 100
attributes(clust_prop)$class <- "matrix"
clust_prop_heatmap <- ComplexHeatmap::Heatmap(matrix = clust_prop,
  col = circlize::colorRamp2(breaks = c(0, 5, 10, 20, 30, 60),
    colors = viridis::inferno(n = 6)), clustering_method_rows = "ward.D2",
  clustering_method_columns = "ward.D2", row_title = "Cluster Identity",
  column_title = "Sample ID", heatmap_legend_param = list(title = "Percent of sample"),
  column_dend_height = unit(15, units = "mm"))
tiff(filename = paste0(results_out, "allMyeloid_cluster-proportion-by-sample_heatmap.tiff"),
  res = 440, height = 6, width = 6, units = "in")
clust_prop_heatmap
invisible(dev.off())
clust_prop_heatmap
```

The cluster dendrogram of the samples shows three larger groupings:

- uninj\_sample1 + uninj\_sample2
- uninj\_sample3 + neonatal datasets
- all other adult SCI datasets

The groupings likely reflect that the adult SCI datasets contain peripherally derived immune cells while the neonatal data had them removed via FACS. Interesting to note that uninj\_sample3 has low proportion of cells from cluster 1, which we described as the homeostatic microglia cluster. This might speak to sample quality. Regardless, because of the neonatal FACS, a better comparison of proportions would be among the microglia clusters.

```
micro_sub <- c(0, 1, 2, 4, 5, 7, 11, 12, 18)
micro_sub <- sci@meta.data[c("integrated_snn_res.0.8", "sample_id")] %>%
  filter(integrated_snn_res.0.8 %in% c(micro_sub))
```

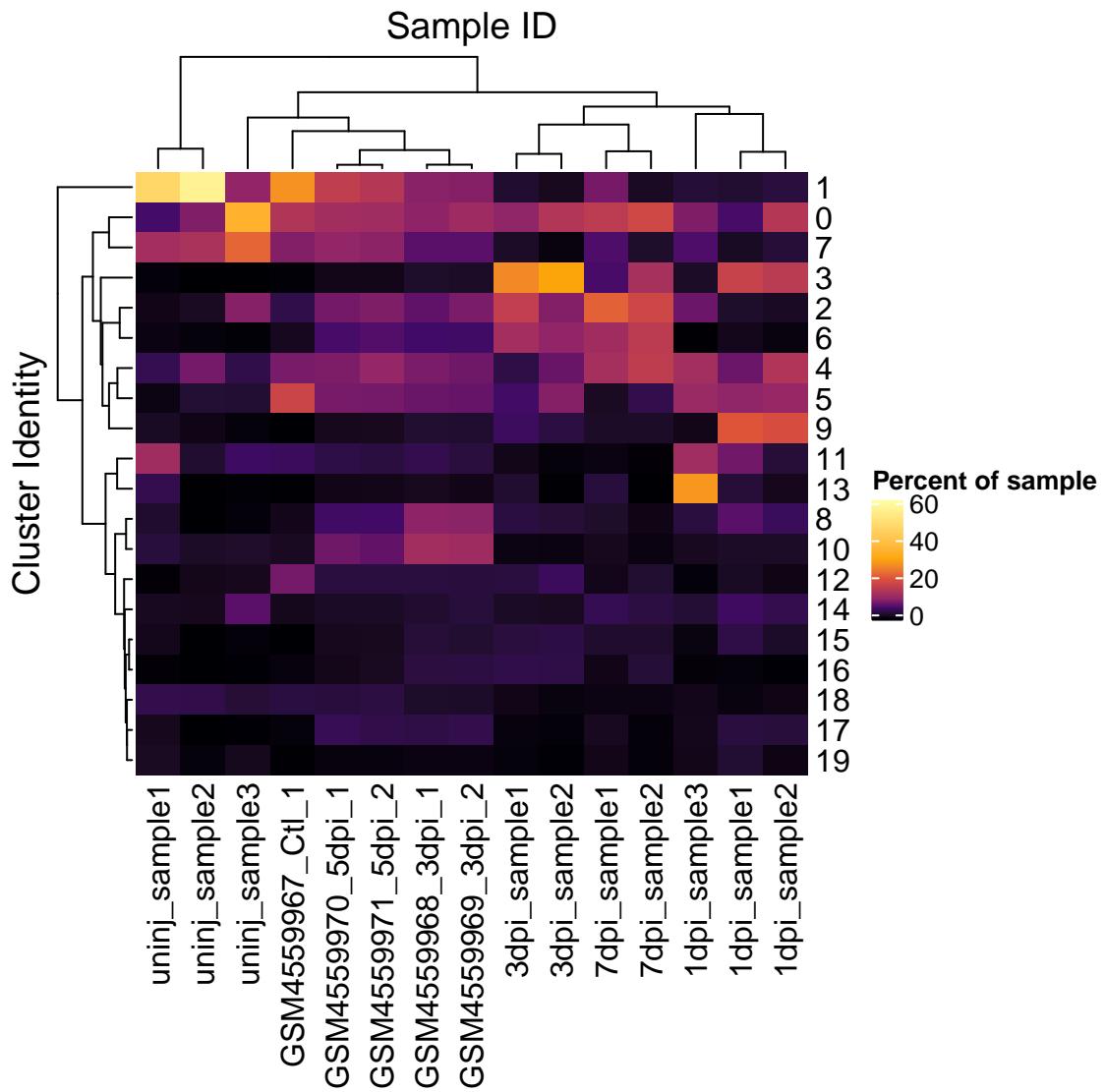


Figure 3: Heatmap of proportion of cells from each cluster by sample.

```

micro_sub$integrated_snn_res.0.8 <- as.character(micro_sub$integrated_snn_res.0.8)
micro_prop <- proportions(x = table(micro_sub$integrated_snn_res.0.8,
  micro_sub$sample_id), margin = 2) * 100
attributes(micro_prop)$class <- "matrix"
micro_prop_heatmap <- ComplexHeatmap::Heatmap(matrix = micro_prop,
  col = circlize::colorRamp2(breaks = c(0, 10, 20, 30, 60),
  colors = viridis::inferno(n = 5)), clustering_method_rows = "ward.D2",
  clustering_method_columns = "ward.D2", row_title = "Cluster Identity",
  column_title = "Sample ID", heatmap_legend_param = list(title = "Percent of sample"),
  column_dend_height = unit(15, units = "mm"))
tiff(filename = paste0(results_out, "microgliaClusters_cluster-proportion-by-sample_heatmap.tiff"),
  res = 440, height = 5, width = 6, units = "in")
micro_prop_heatmap
invisible(dev.off())
micro_prop_heatmap

```

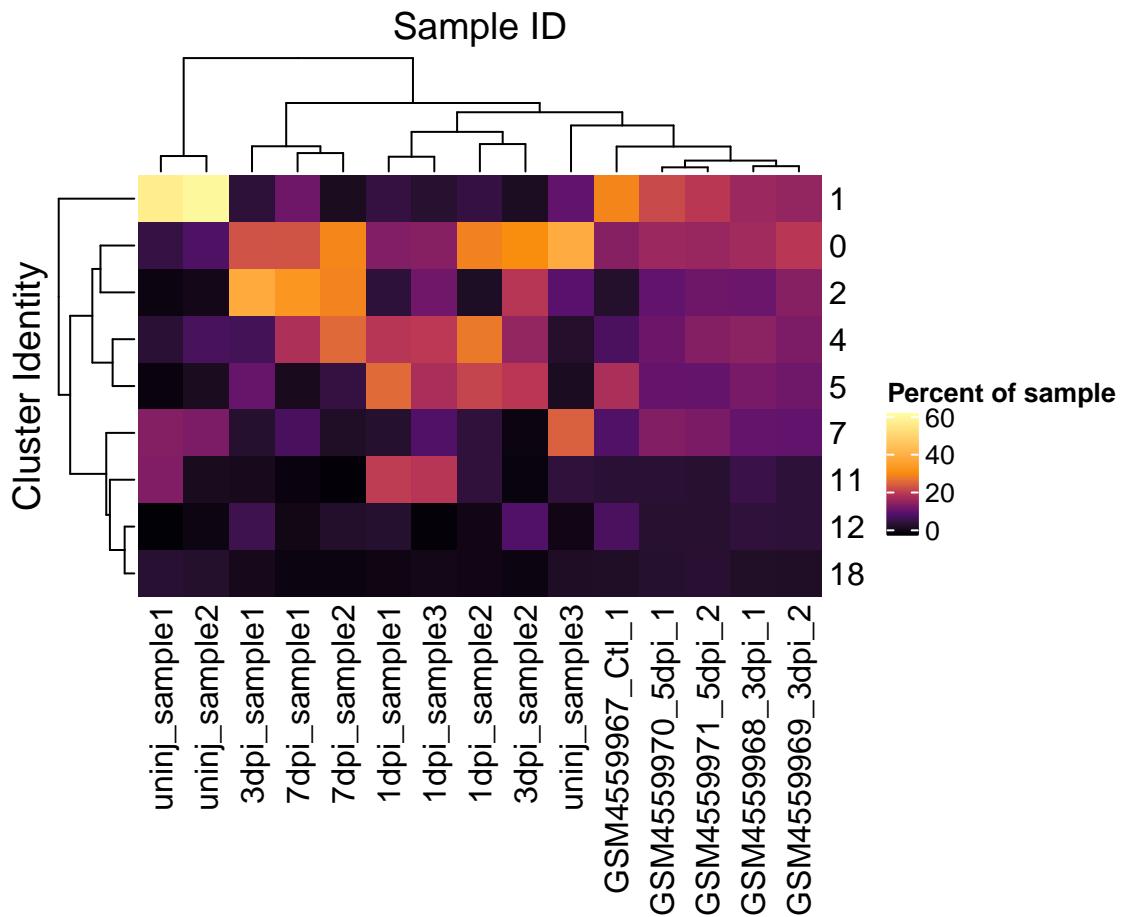


Figure 4: Heatmap of proportion of cells from each microglia cluster by sample.

```

# saveRDS(sci, file = '../data/myeloid_combined.rds')

sessionInfo()

```

```

## R version 4.0.2 (2020-06-22)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19041)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252 LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252 LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats      graphics   grDevices utils      datasets   methods    base
##
## other attached packages:
## [1] cowplot_1.1.0 dplyr_1.0.2   ggplot2_3.3.2 Seurat_3.2.2
##
## loaded via a namespace (and not attached):
## [1] Rtsne_0.15           colorspace_1.4-1     rjson_0.2.20       deldir_0.1-29
## [5] ellipsis_0.3.1       ggridges_0.5.2       circlize_0.4.10    GlobalOptions_0.1.2
## [9] clue_0.3-57          spatstat.data_1.4-3  leiden_0.3.3       listenv_0.8.0
## [13] farver_2.0.3         ggrepel_0.8.2       fansi_0.4.1        codetools_0.2-16
## [17] splines_4.0.2        knitr_1.30          polyclip_1.10-0    jsonlite_1.7.1
## [21] ica_1.0-2            cluster_2.1.0       png_0.1-7          uwot_0.1.8
## [25] shiny_1.5.0          sctransform_0.3     compiler_4.0.2     httr_1.4.2
## [29] assertthat_0.2.1     Matrix_1.2-18       fastmap_1.0.1      lazyeval_0.2.2
## [33] cli_2.0.2             later_1.1.0.1      formatR_1.7        htmltools_0.5.0
## [37] tools_4.0.2           rsvd_1.0.3          igraph_1.2.5       gtable_0.3.0
## [41] glue_1.4.2            RANN_2.6.1          reshape2_1.4.4     rappdirs_0.3.1
## [45] tinytex_0.26          Rcpp_1.0.5          spatstat_1.64-1    vctrs_0.3.4
## [49] nlme_3.1-148          lmtest_0.9-38      xfun_0.18          stringr_1.4.0
## [53] globals_0.13.0        mime_0.9            miniUI_0.1.1.1    lifecycle_0.2.0
## [57] irlba_2.3.3          goftest_1.2-2       future_1.19.1     MASS_7.3-51.6
## [61] zoo_1.8-8              scales_1.1.1        promises_1.1.1    spatstat.utils_1.17-0
## [65] parallel_4.0.2         RColorBrewer_1.1-2  ComplexHeatmap_2.4.3 yaml_2.2.1
## [69] reticulate_1.16        pbapply_1.4-3       gridExtra_2.3      rpart_4.1-15
## [73] stringi_1.5.3         highr_0.8           shape_1.4.5        rlang_0.4.7
## [77] pkgconfig_2.0.3        matrixStats_0.57.0  evaluate_0.14     lattice_0.20-41
## [81] ROCR_1.0-11           purrr_0.3.4         tensor_1.5        patchwork_1.0.1
## [85] htmlwidgets_1.5.2      labeling_0.3        tidyselect_1.1.0   RcppAnnoy_0.0.16
## [89] plyr_1.8.6              magrittr_1.5        R6_2.4.1           generics_0.0.2
## [93] pillar_1.4.6            withr_2.3.0         mgcv_1.8-31       fitdistrplus_1.1-1
## [97] survival_3.1-12        abind_1.4-5         tibble_3.0.3       future.apply_1.6.0
## [101] crayon_1.3.4           KernSmooth_2.23-17  plotly_4.9.2.1    rmarkdown_2.4
## [105] viridis_0.5.1          GetoptLong_1.0.4     grid_4.0.2         data.table_1.13.0
## [109] digest_0.6.25          xtable_1.8-4        tidyverse_1.1.2    httpuv_1.5.4
## [113] munsell_0.5.0          viridisLite_0.3.0

```