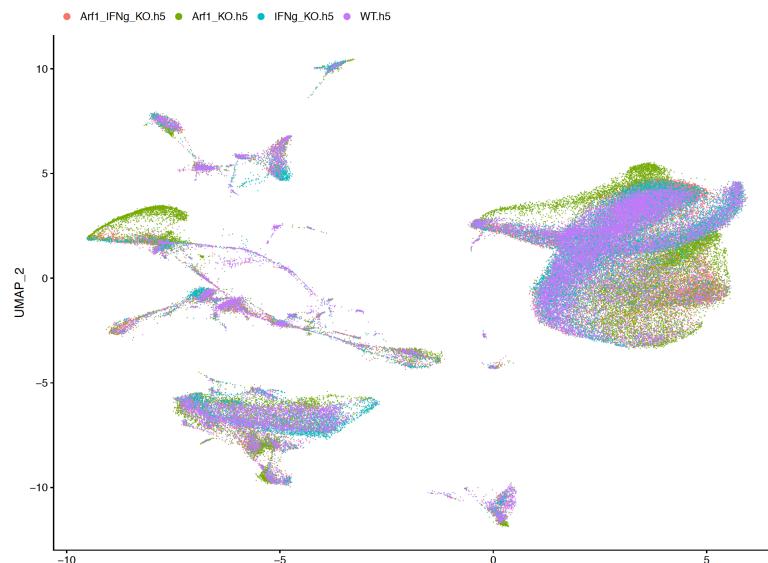


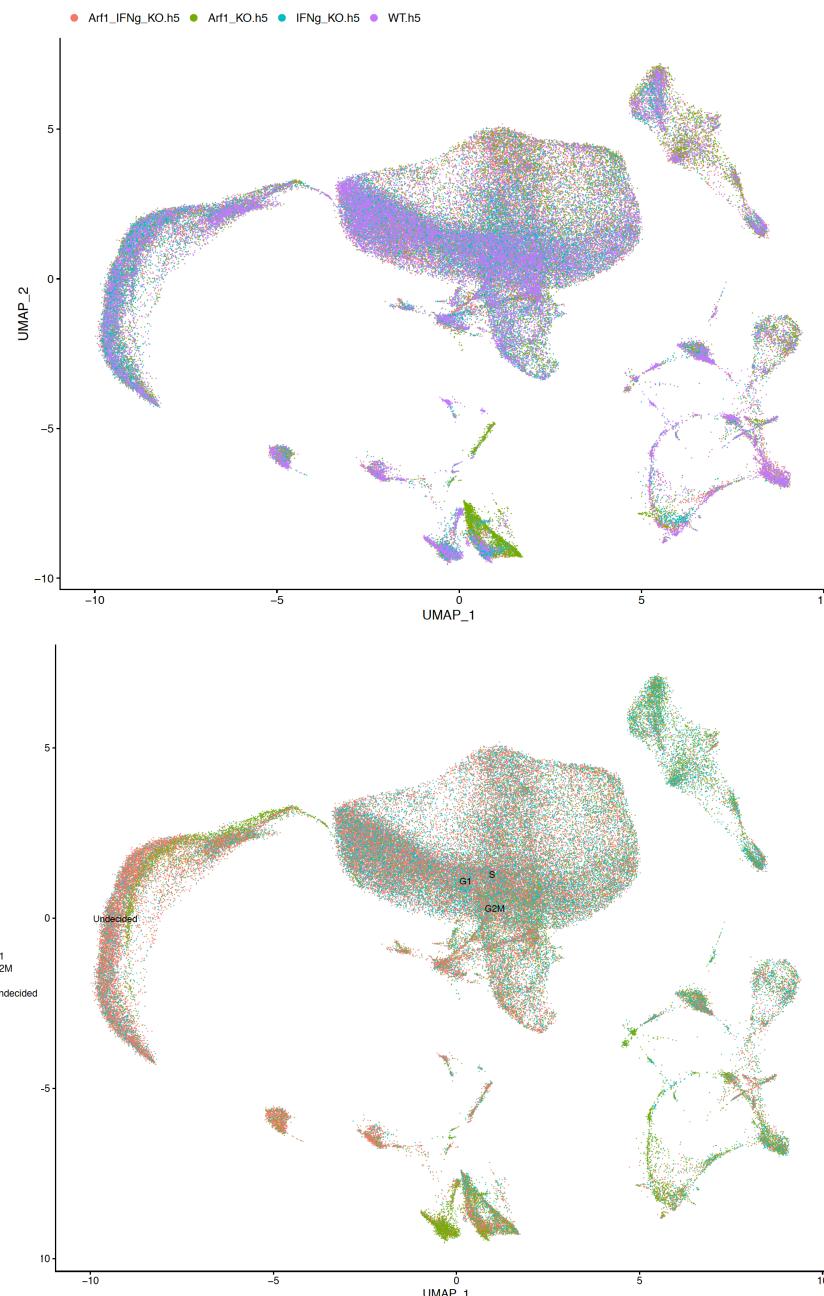
# QC and analysis of four scRNA samples

Ccbr 1045

## Before batch correction



## Batch-corrected



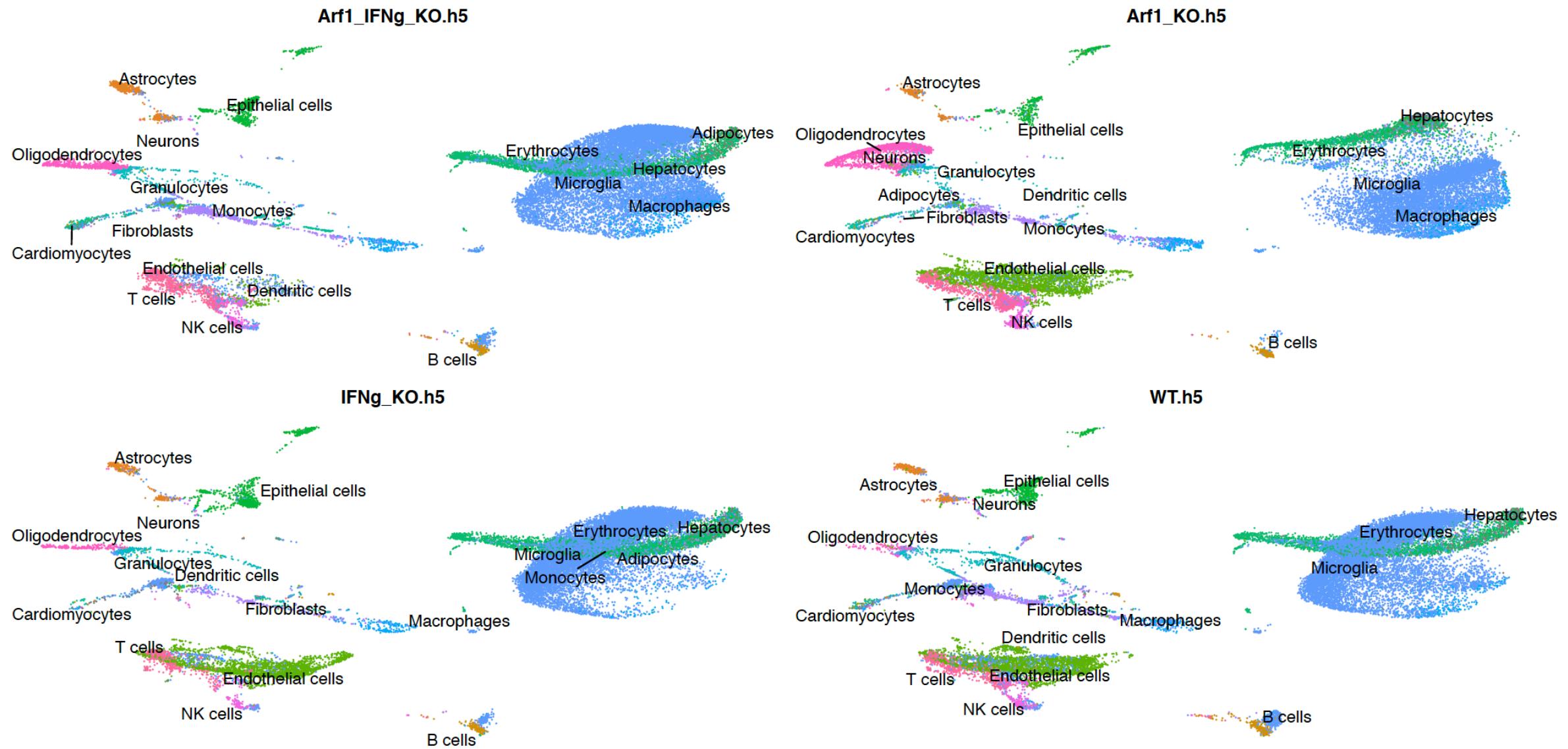
### Color by sample

Plotting all samples overlaid, colored by sample. Check to see if clustering behavior is tied to sample.  
It looks like the four samples aren't separated before or after, which is expected.

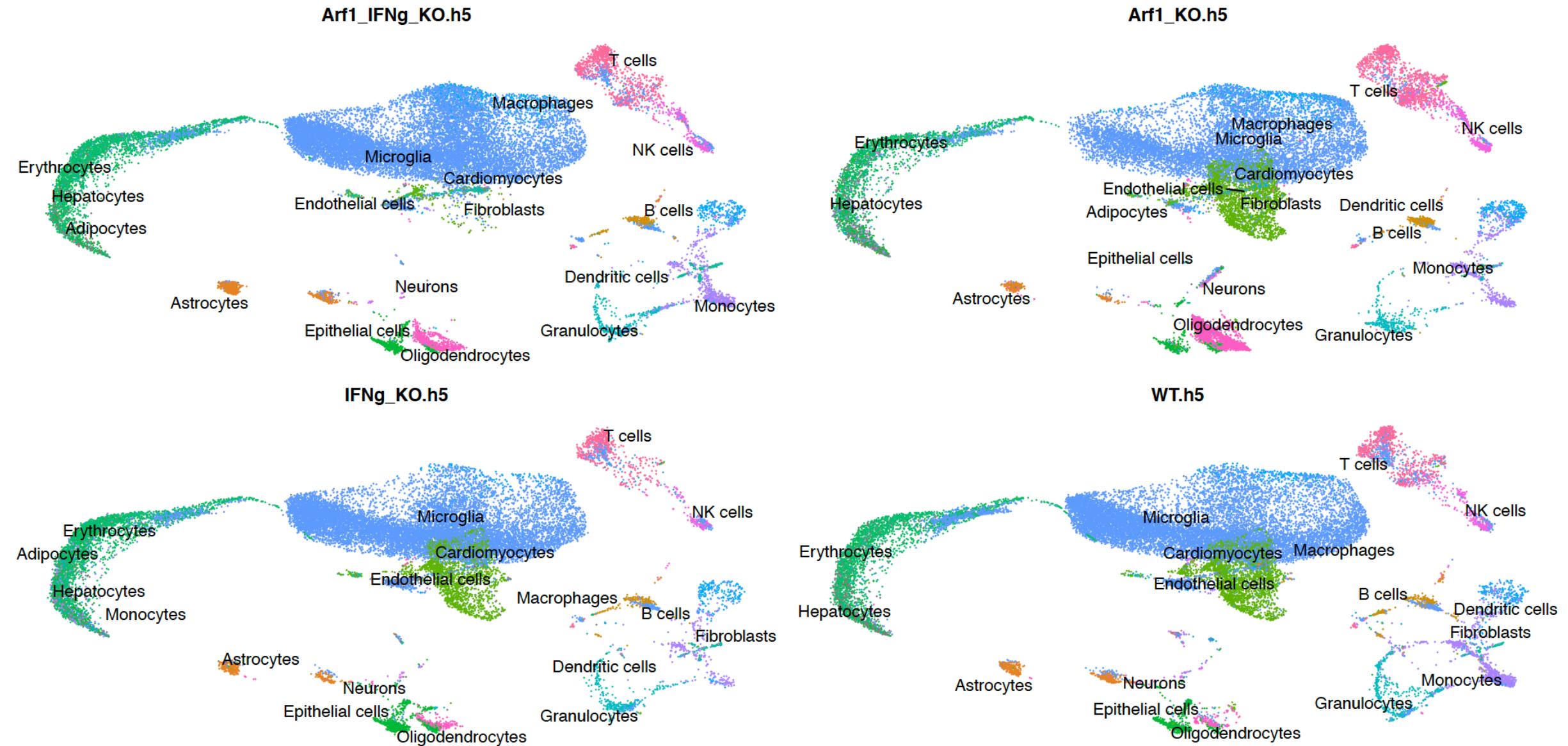
### Color by cell cycle

Plotting all samples overlaid, colored by phase. Check to see if clustering behavior is tied to cell cycle.  
It looks like the microglial cluster(main cluster) isn't tied to cell cycle, good.

## Before batch correction, group by mouseRNAseq\_main

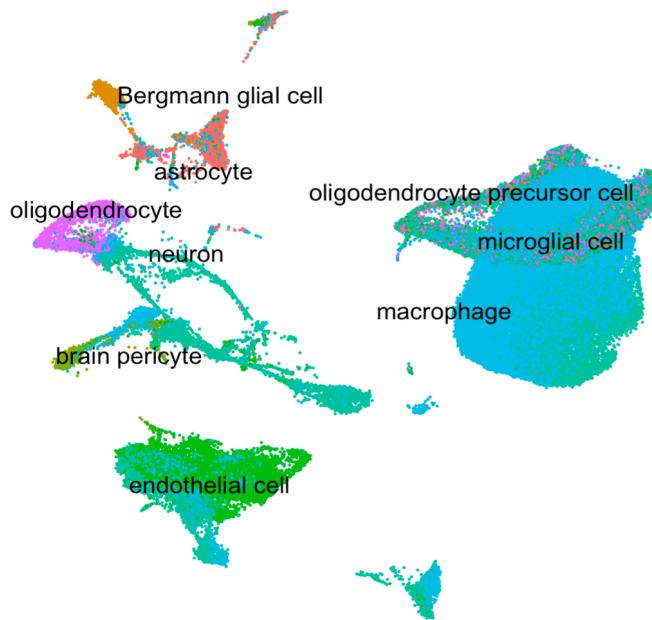


# Batch-corrected, group by mouseRNAseq\_main

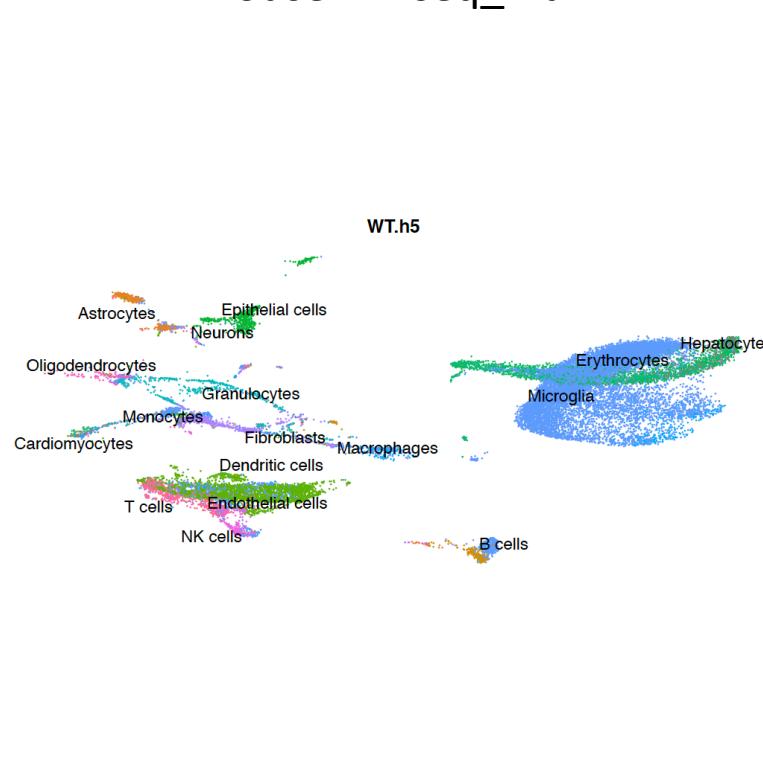


# Comparison of cell identity annotation by three databases

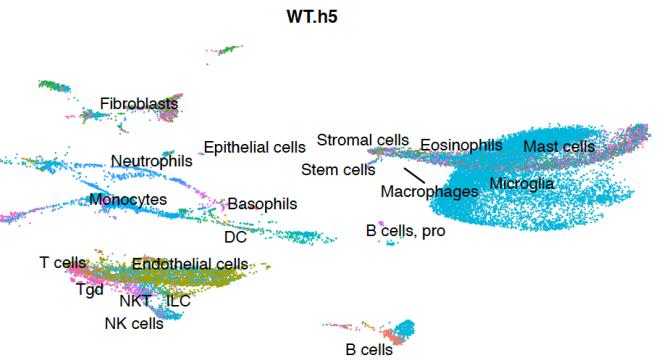
Tabulus Muris



mouseRNAseq\_main

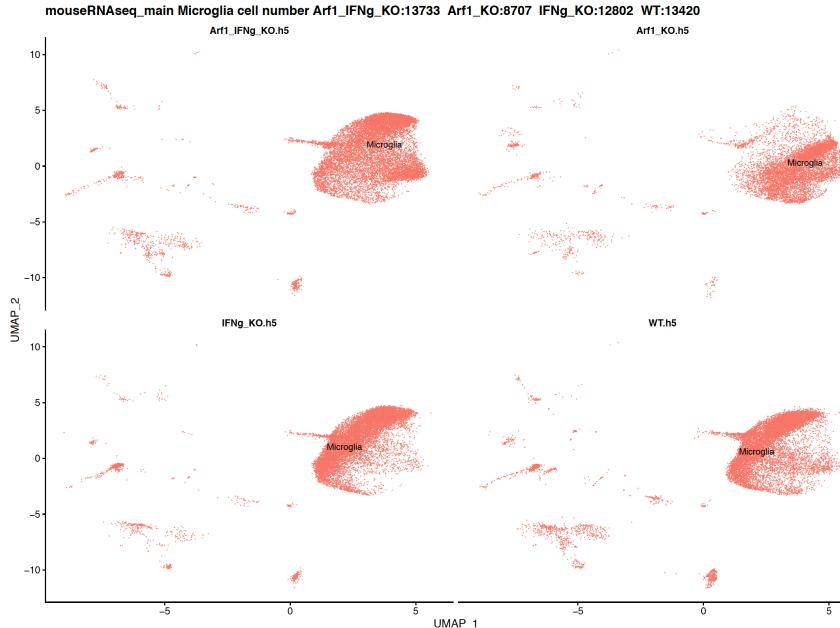
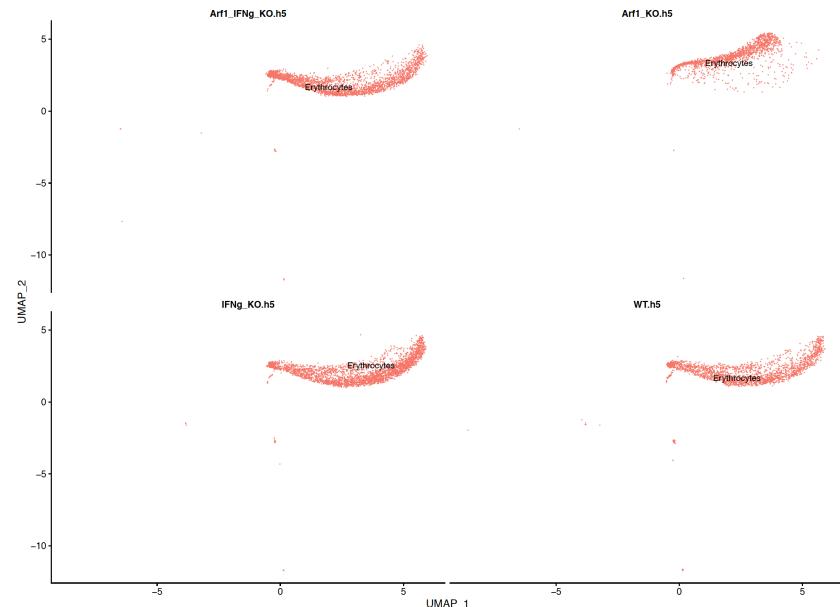


Immgen\_main



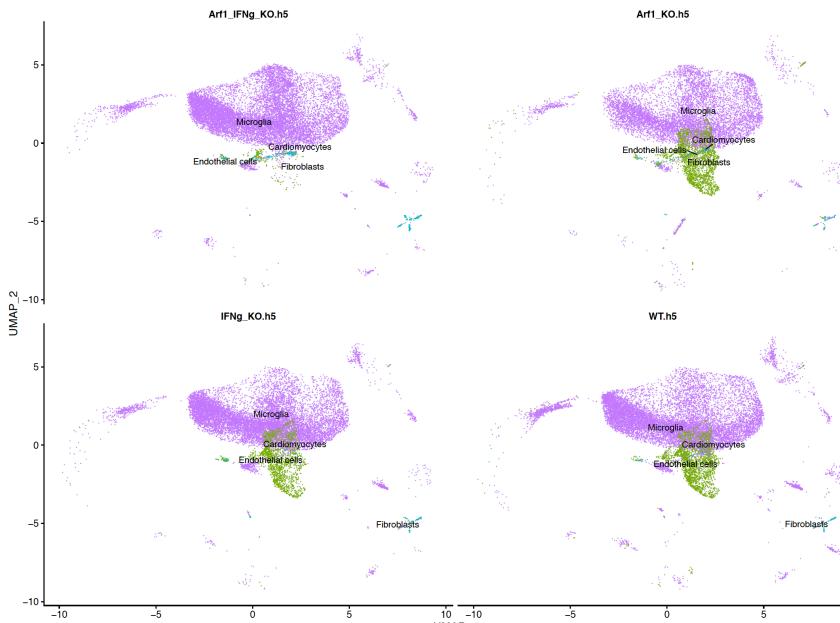
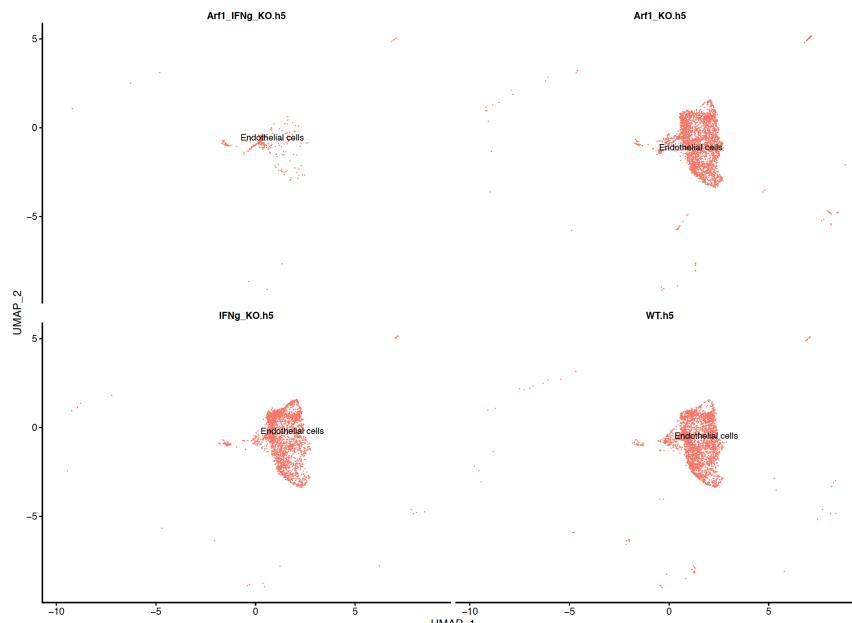
First, I checked to see which database is more accurate in assigning cell type by using FindMarkers. This is UMAP showing WT sample before batch correction. For cluster7, 9, 10,11, the top enriched genes are Hbb-bt, Hba-a1, Hba-a2, etc. all hemoglobin genes, indicating the cells in cluster7, 9, 10 represent red blood cells. This is accurately predicted by the mouseRNAseq\_main as erythrocytes. In addition, mouseRNAseq\_main also cluster23 as astrocytes correctly as the top enriched gene Aldoc is an astrocyte marker. Also mouseRNAseq\_main predicted cluster15 as Oligodendrocytes, which is has high Plp1 expression(Oligodendrocytes marker).

# Plotting individual cells to see their position



Before batch-correction

Erythrocytes as labelled by mouseRNAseq database, cluster with Microglia

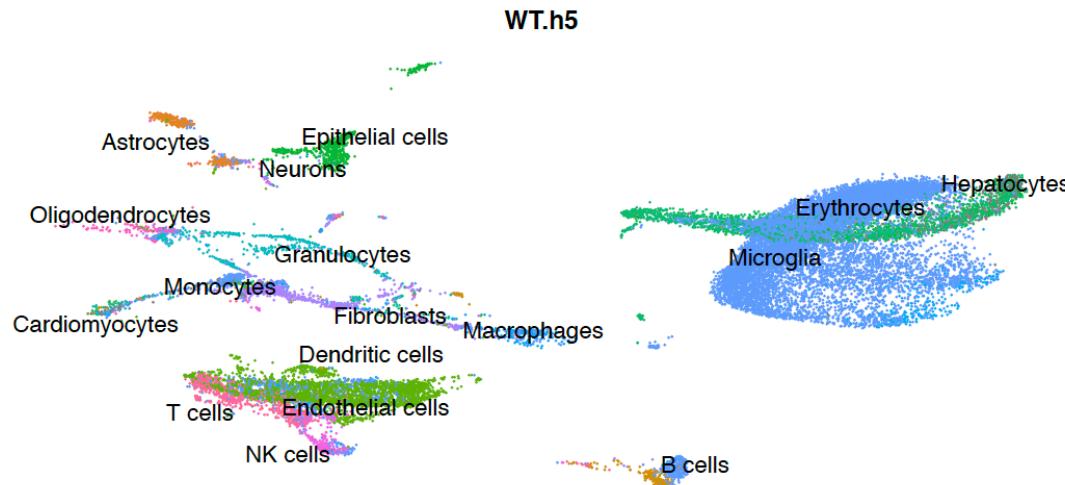


After batch-correction

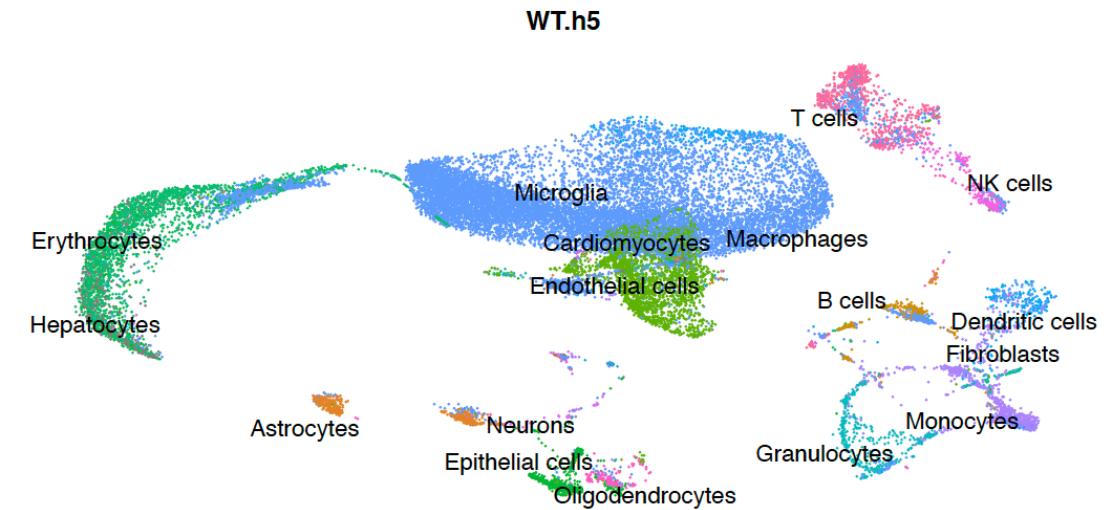
Endothelial cells as labelled by mouseRNAseq database, cluster with Microglia

Batch-corrected sample seems to have better cluster separation?

Before batch correction



Batch-corrected

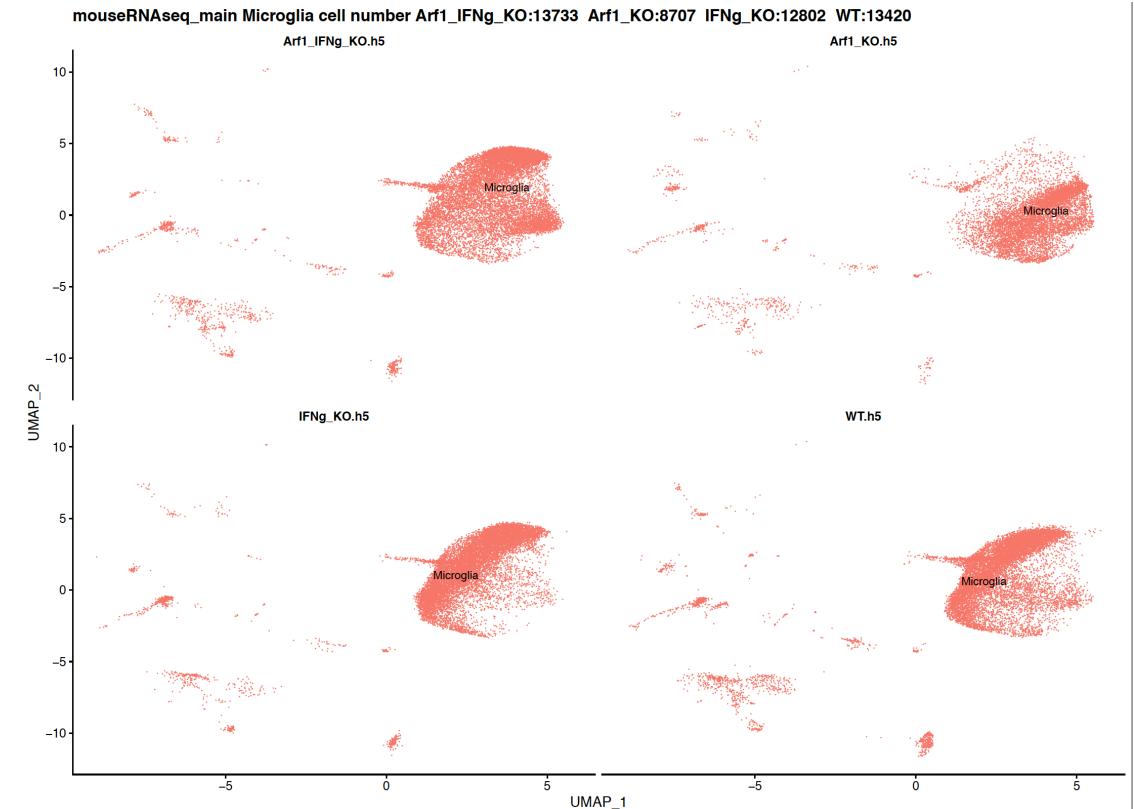
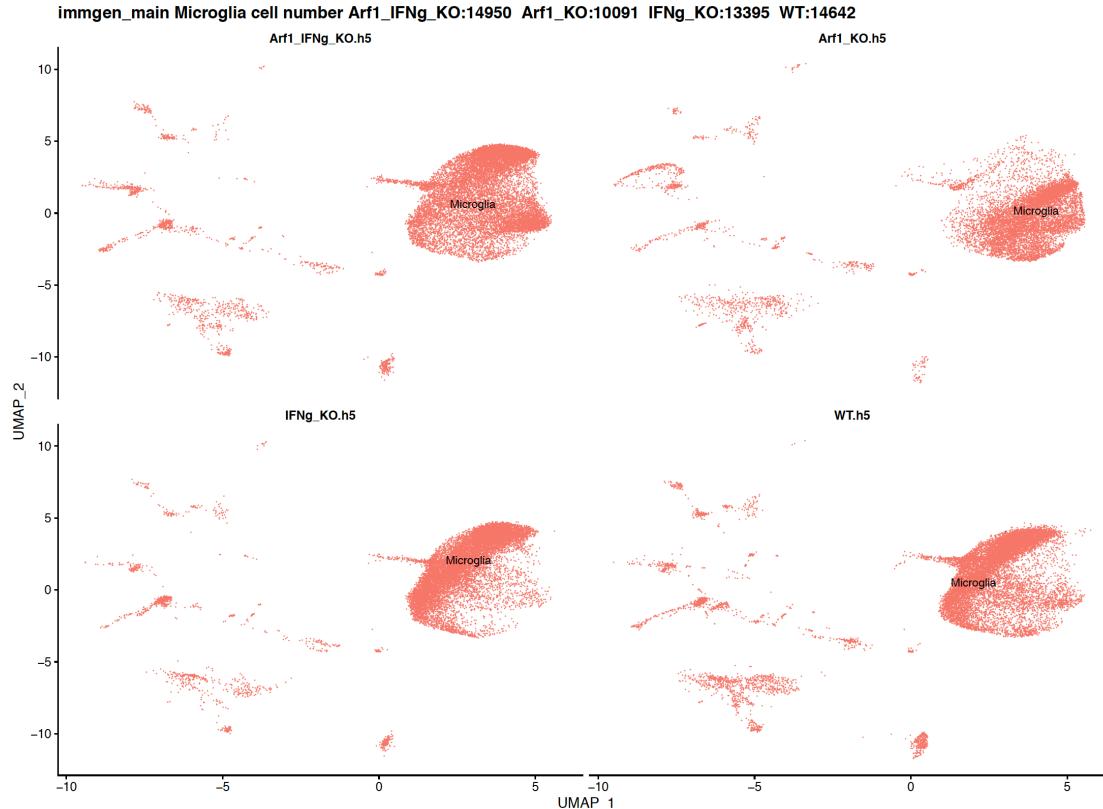


Is before batch correction or after more of how we expect the clusters to separate? Before batch correction, microglia cell clusters are mixed with erythrocytes, which is not expected. Also Endothelial cells and T cells are mixed, which is not expected either. After batch correction, there seems to be better separation of distinct cell types, for example, erythrocytes, microglia, and T cells are well separated. Although the endothelial cells got mixed with microglia cluster. The other concern I have with batch corrected is the microglia cluster becomes very homogenized (see Page3, overlaid samples), it is possible some biologically interesting variation may be removed too.

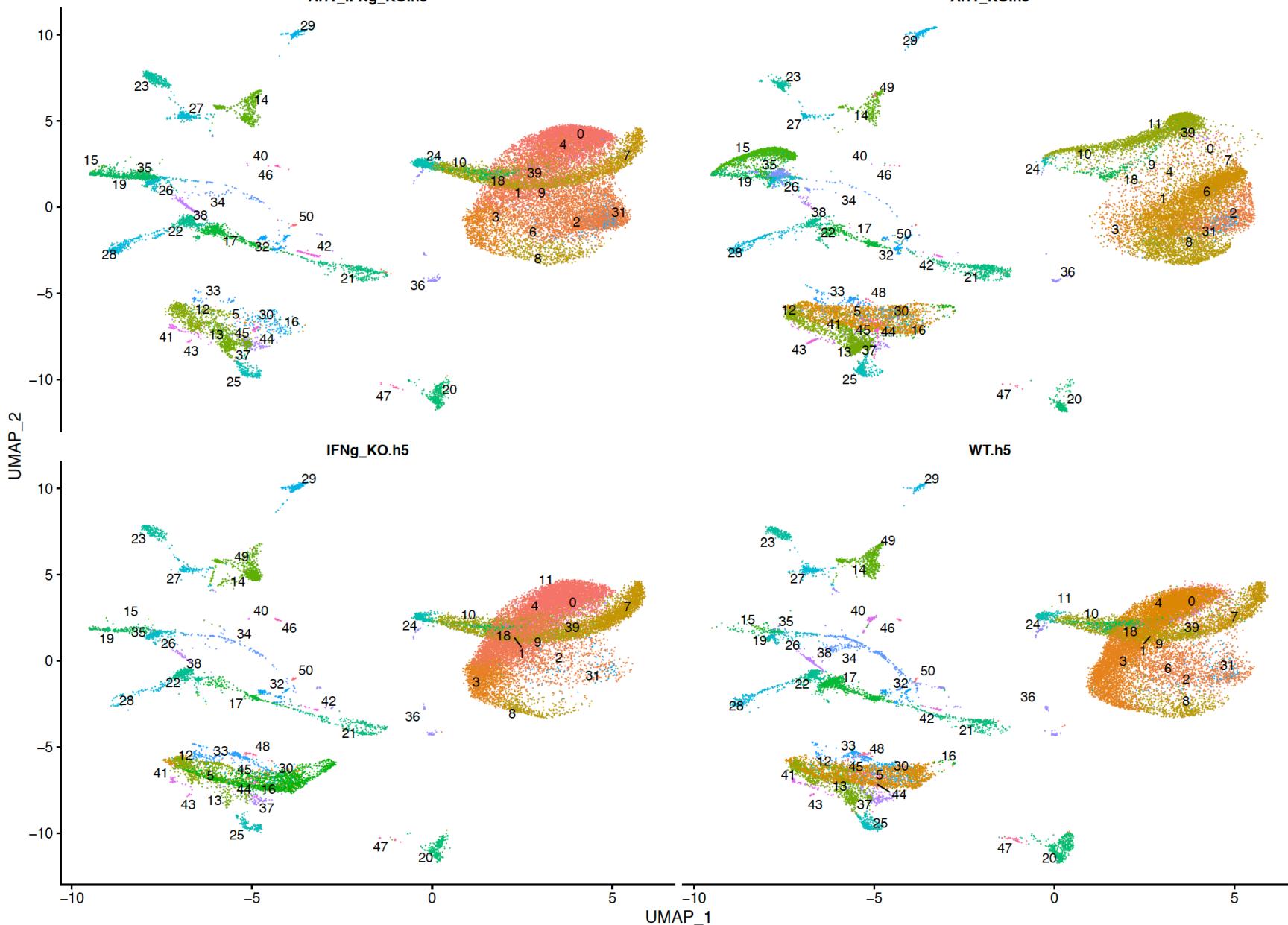
There isn't too much changing before and after batch correction regarding the samples themselves. So we might go with before batch-correction. Just a few more checks...

# Checking to see the number of microglial cells

The number of cells labeled as microglia by immgen and mouseRNAseq.  
Since the study focus is microglia, we should expect good concordance between the two.  
Immgene has ~10% more cells labelled as microglia.

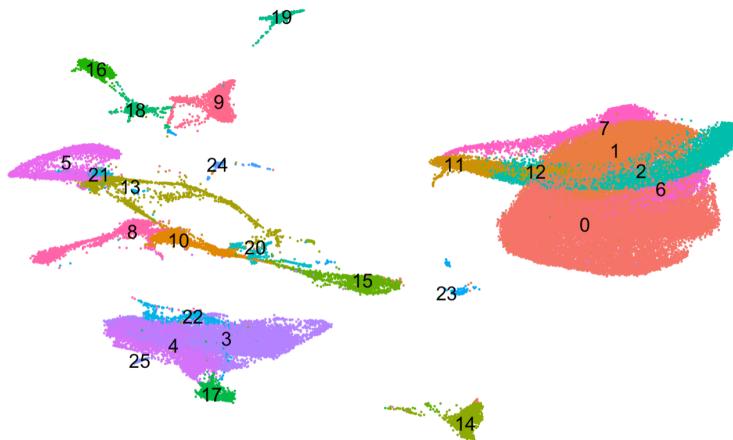


# Seurat clusters res1.2 split by sample

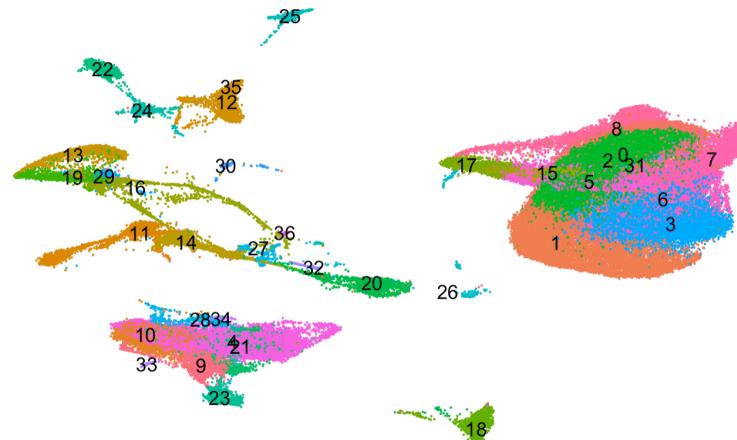


seurat_clu_sters	mouseRNase_q_main
0	Microglia
1	Microglia
2	Microglia
3	Microglia
4	Microglia
5	NA
6	Microglia
7	NA
8	Microglia
9	Erythrocytes
10	Erythrocytes
11	Erythrocytes
12	T cells
13	NK cells
14	Epithelial cells
15	Oligodendrocytes
16	Microglia
17	Monocytes
18	Microglia
19	Oligodendrocytes
20	Microglia
21	Macrophages
22	Microglia
23	Astrocytes
24	Erythrocytes
25	NK cells
seurat_cluste	mouseRNaseq_main
26	Granulocytes
27	Neurons
28	Monocytes
29	Epithelial cells
30	Microglia
31	Microglia
32	Fibroblasts
33	Endothelial cells
34	Granulocytes
35	Neurons
36	Microglia
37	Endothelial cells
38	Monocytes
39	Epithelial cells
40	Microglia
41	T cells
42	Fibroblasts
43	Microglia
44	T cells
45	Endothelial cells
46	Fibroblasts
47	T cells
48	Endothelial cells
49	Epithelial cells
50	Monocytes

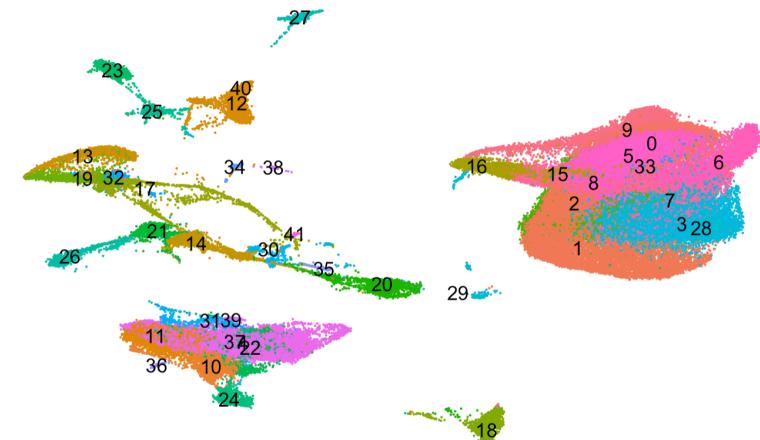
Resolution 0.3



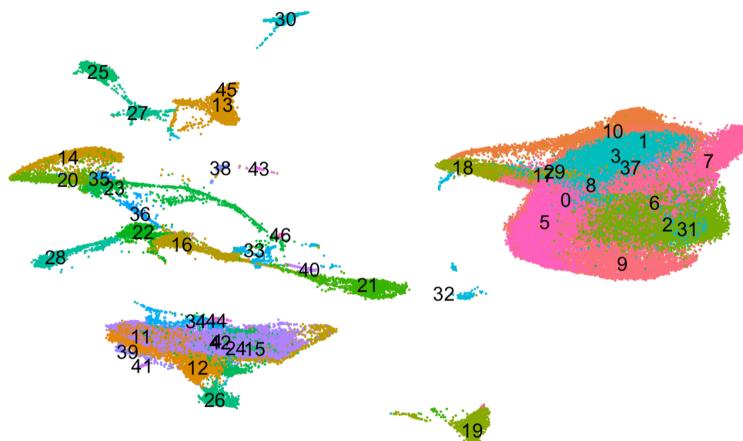
Resolution 0.6



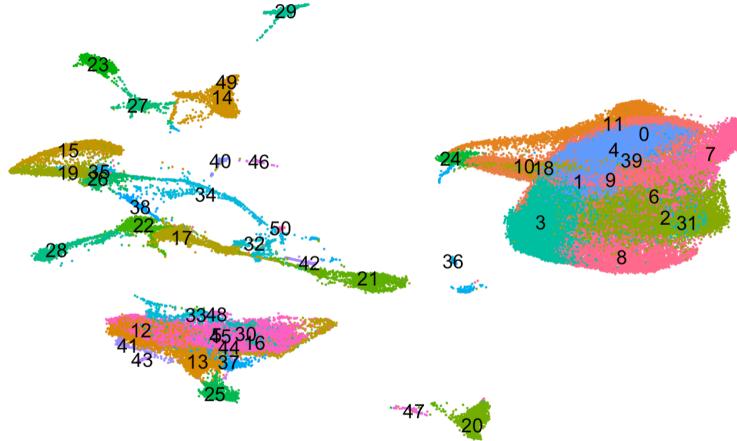
Resolution 0.8



Resolution 1



Resolution 1.2



## Cell population numbers annotated by mouseRNAseq\_main

Microglia	Oligodendrocytes	Erythrocytes	Astrocytes	Epithelial cells	NK cells	T cells
48662	3601	10068	1556	4188	1403	3352
Macrophages	Granulocytes	Fibroblasts	Endothelial cells	Neurons	Monocytes	B cells
3339	1485	1023	8127	224	2883	1046
Hepatocytes	Cardiomyocytes	Dendritic cells	Adipocytes			
117	38	39	11			

# DE objectives

- Confirm which clusters are microglia, features=c("Cx3cr1", "Tmem119", "Aif1", "P2ry12", "Cd68", "Cd45").
- Confirm which clusters that mix with microglia aren't microglia

## DEG analysis questions:

Within combined sample, microglia clusters: (to find markers for each cluster)

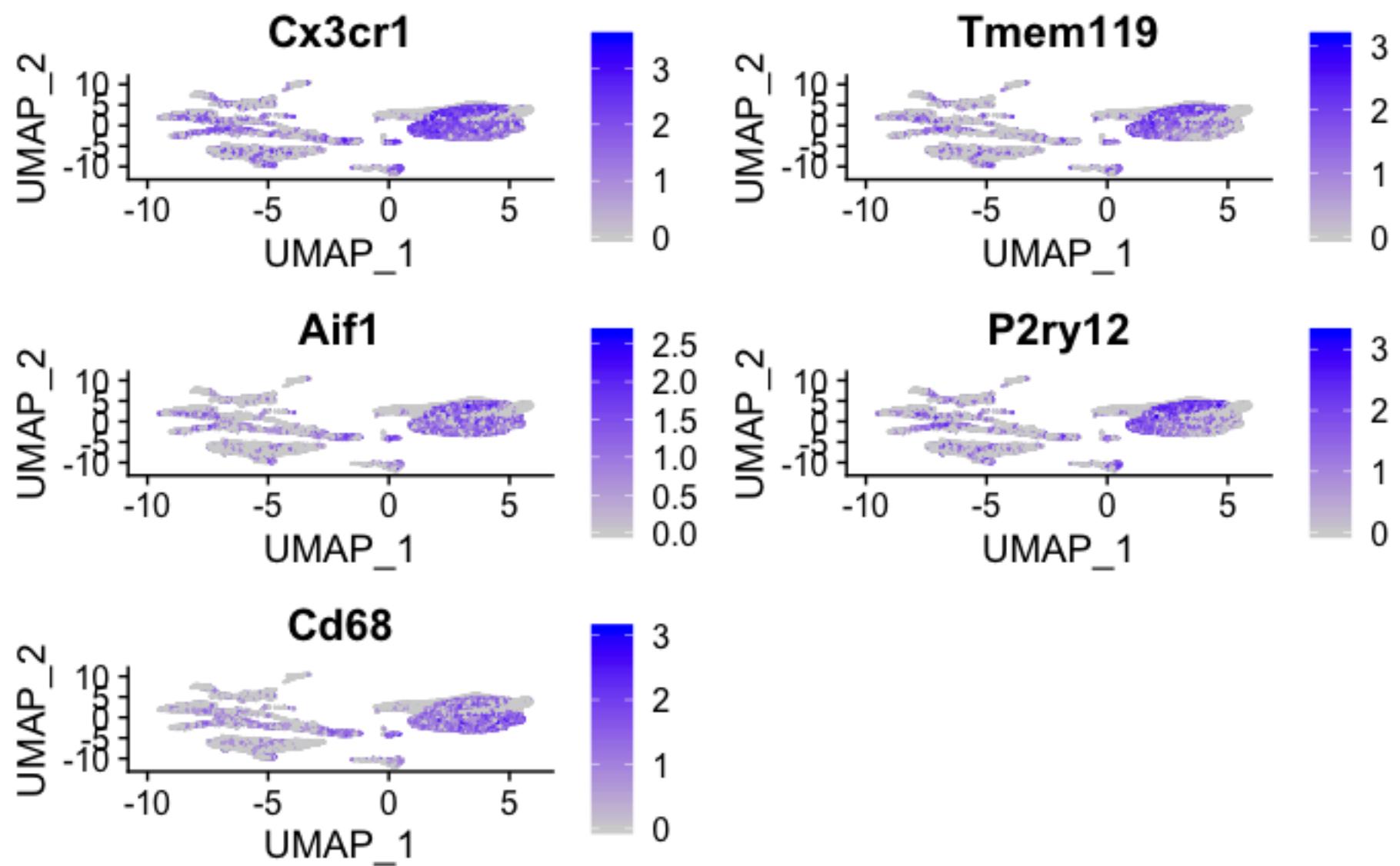
1. Compare cluster 0 to other clusters; cluster 1 to other clusters;

Between sample, microglia clusters: (to find markers for the Arf1 sample)

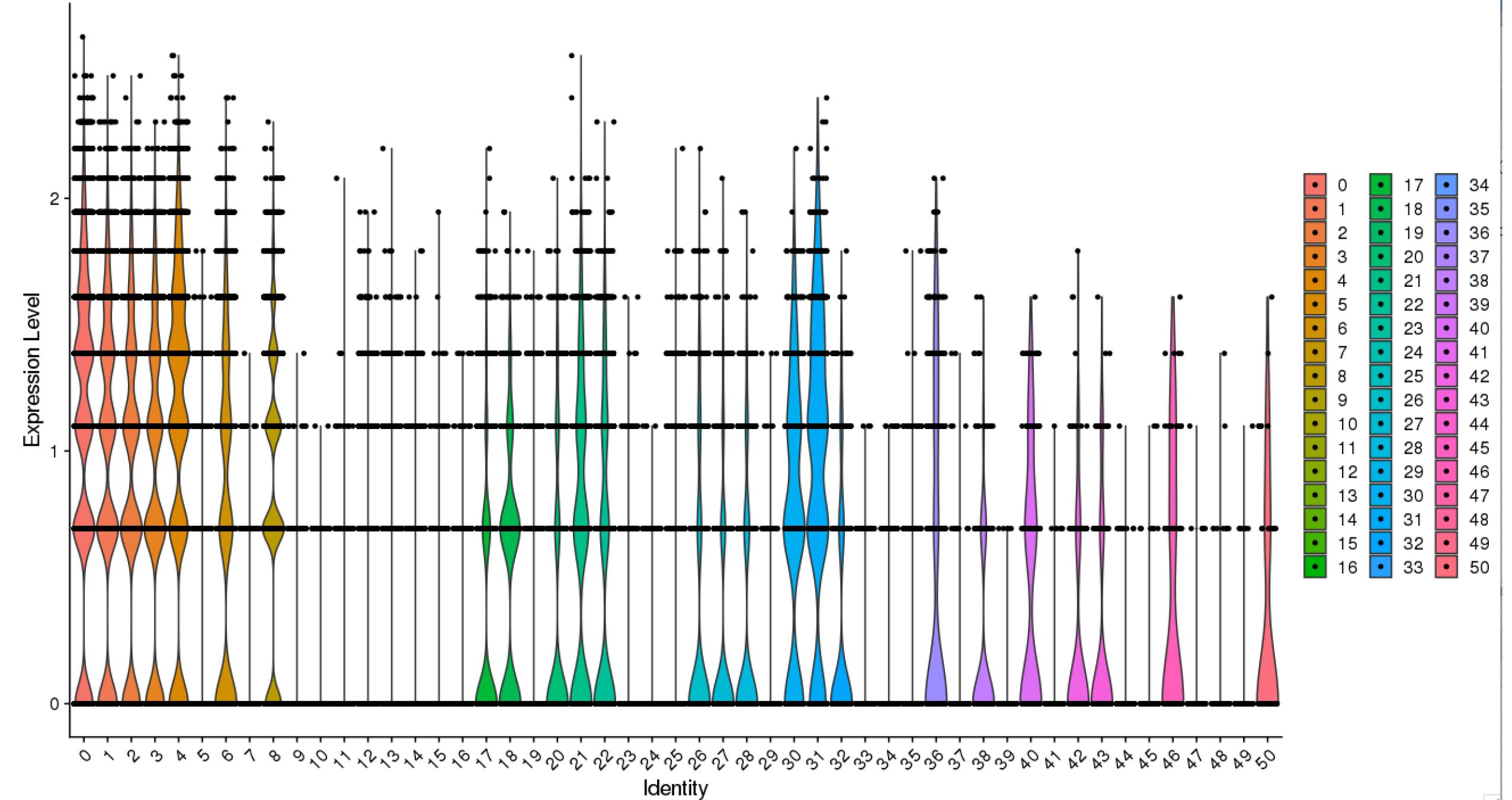
- Compare each knock-down sample to WT (treating WT as control)
  1. Arf1-WT(cluster0), Arf1\_IFNg-WT(cluster0), IFNg-WT(cluster0);
  2. Arf1-WT(cluster1), Arf1\_IFNg-WT(cluster1), IFNg-WT(cluster1);
  3. Arf1-WT(cluster0, 1, 2, 3, 4, 6, 8), Arf1\_IFNg-WT(cluster0, 1, 2, 3, 4, 6, 8), IFNg-WT(cluster0, 1, 2, 3, 4, 6, 8);
  4. Arf1-WT("Microglia"), Arf1\_IFNg-WT("Microglia"), IFNg-WT ("Microglia")
- Compare Arf1 sample to each other sample? (treating Arf1 as control)
  1. Arf1-WT(cluster0), Arf1-Arf1\_IFNg (cluster0), Arf1-IFNg(cluster0);

Compare all the microglia clusters (annotated by mouseRNAseq\_main) to other clusters and test for differential expression of the microglia marker genes.

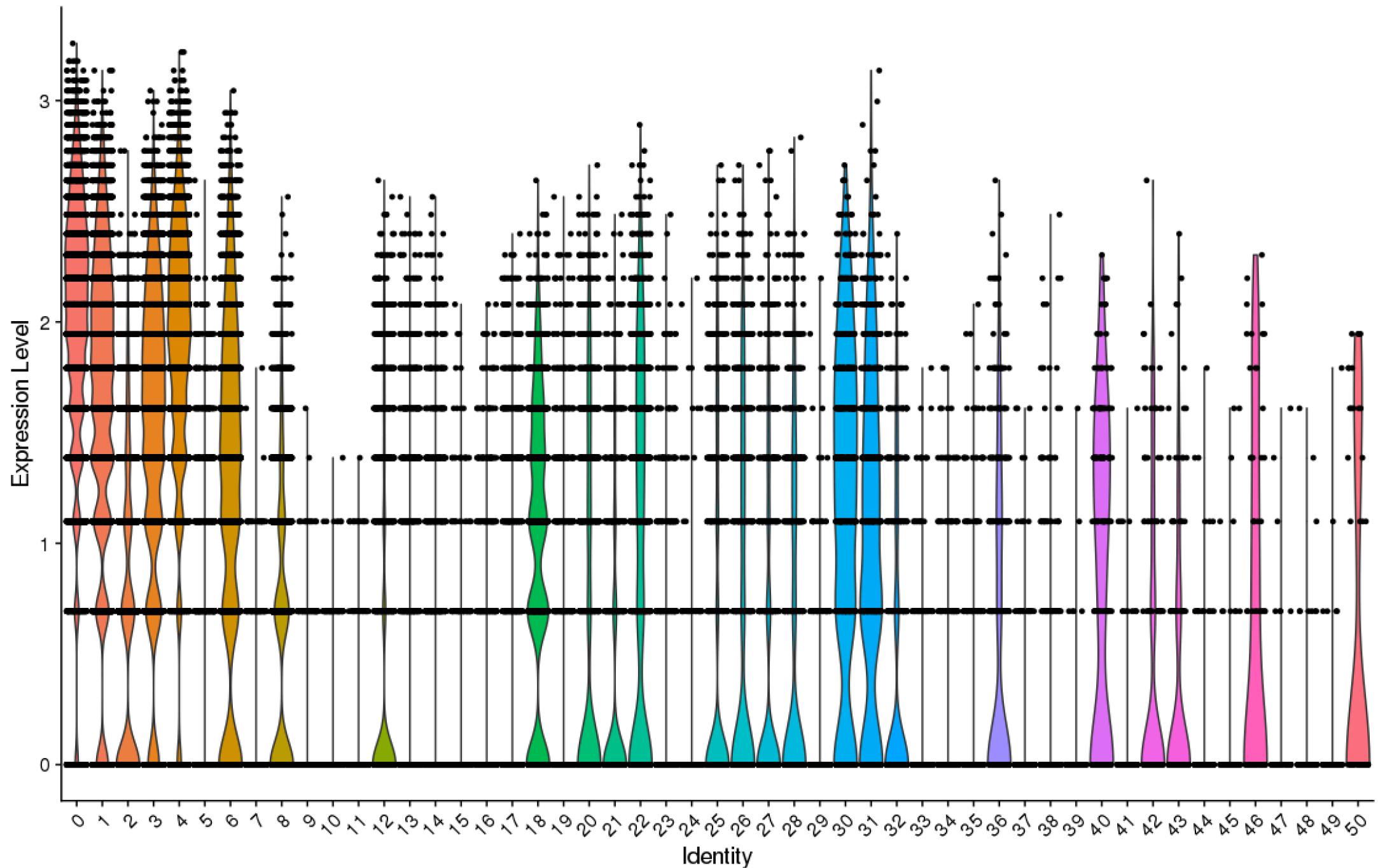
	p_val	avg_logFC	pct.1	pct.2	p_val_adj
<b>Cx3cr1</b>	0	1.6379455	0.977	0.213	0
<b>P2ry12</b>	0	1.4957834	0.871	0.123	0
<b>Tmem119</b>	0	1.4170923	0.869	0.112	0
<b>Aif1</b>	0	0.8432612	0.761	0.132	0
<b>Cd68</b>	0	0.7927106	0.772	0.166	0



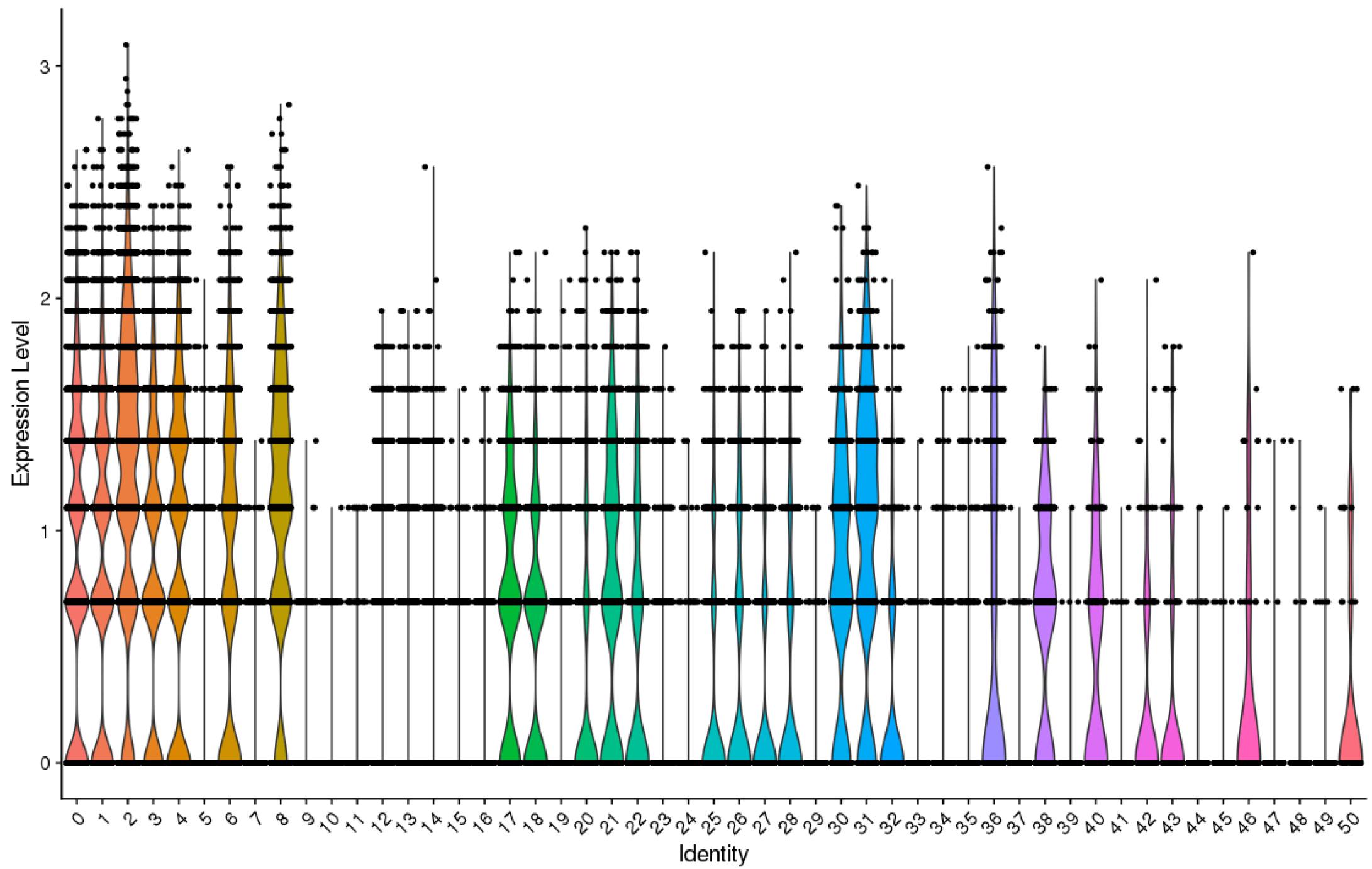
# Aif1

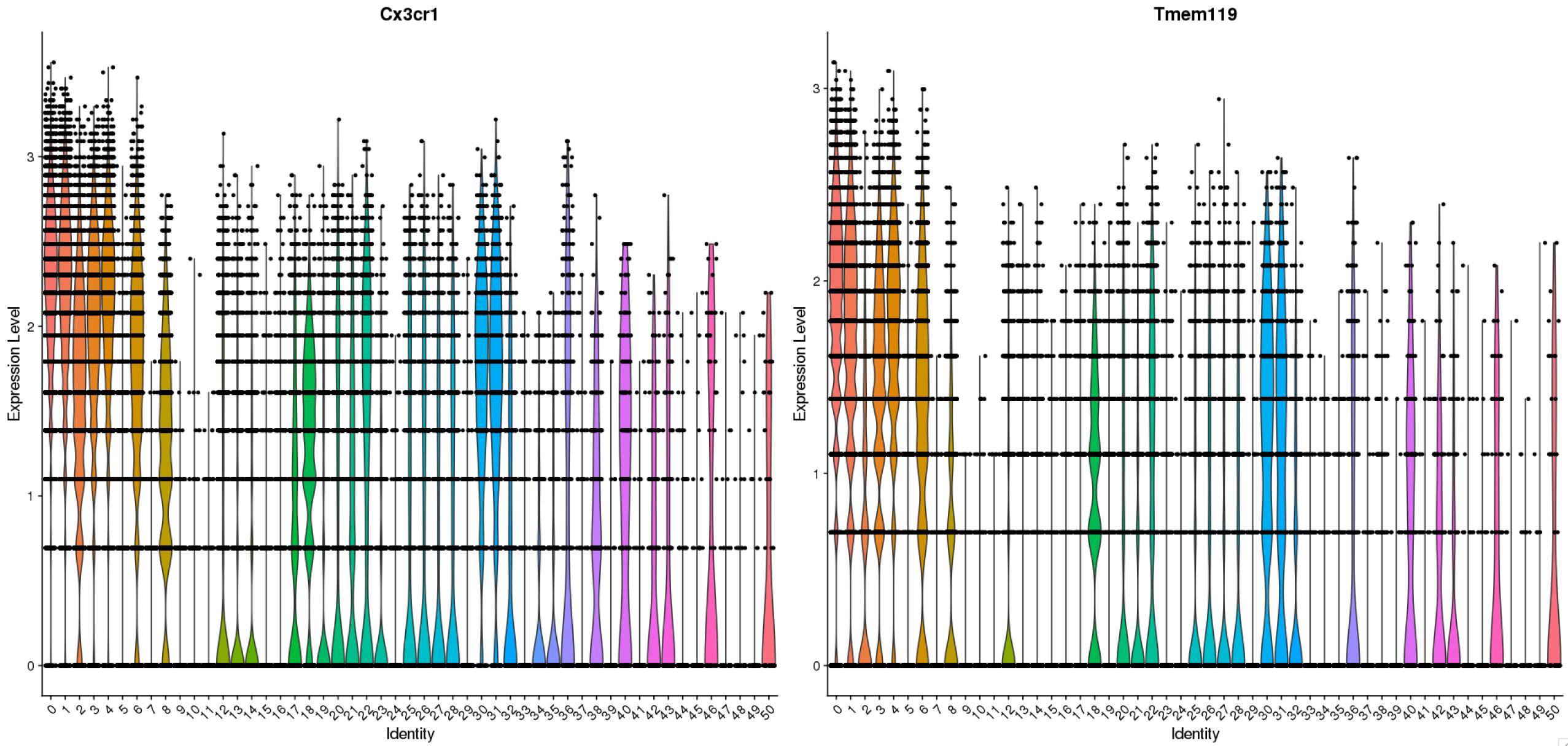


# P2ry12

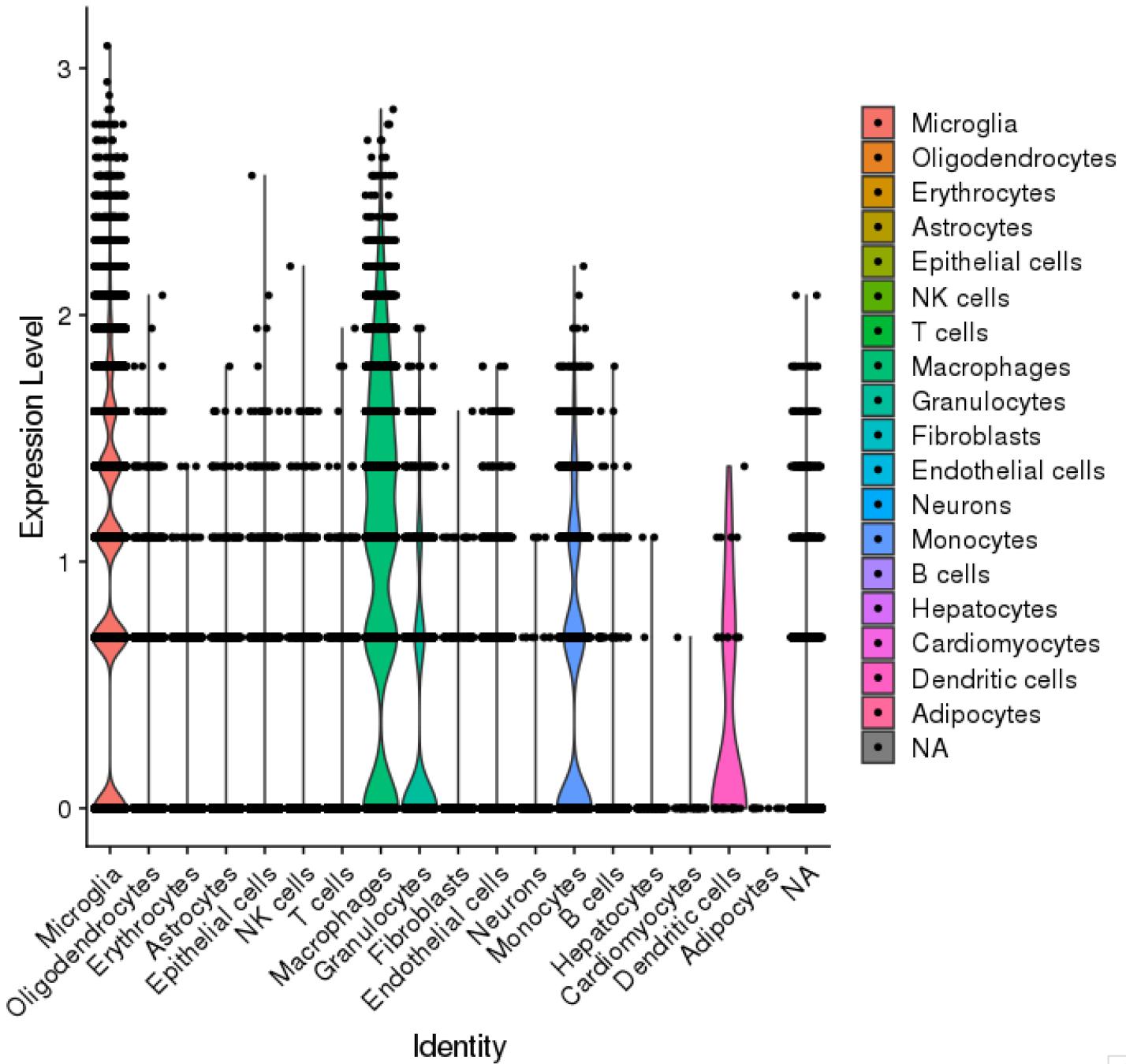


# Cd68

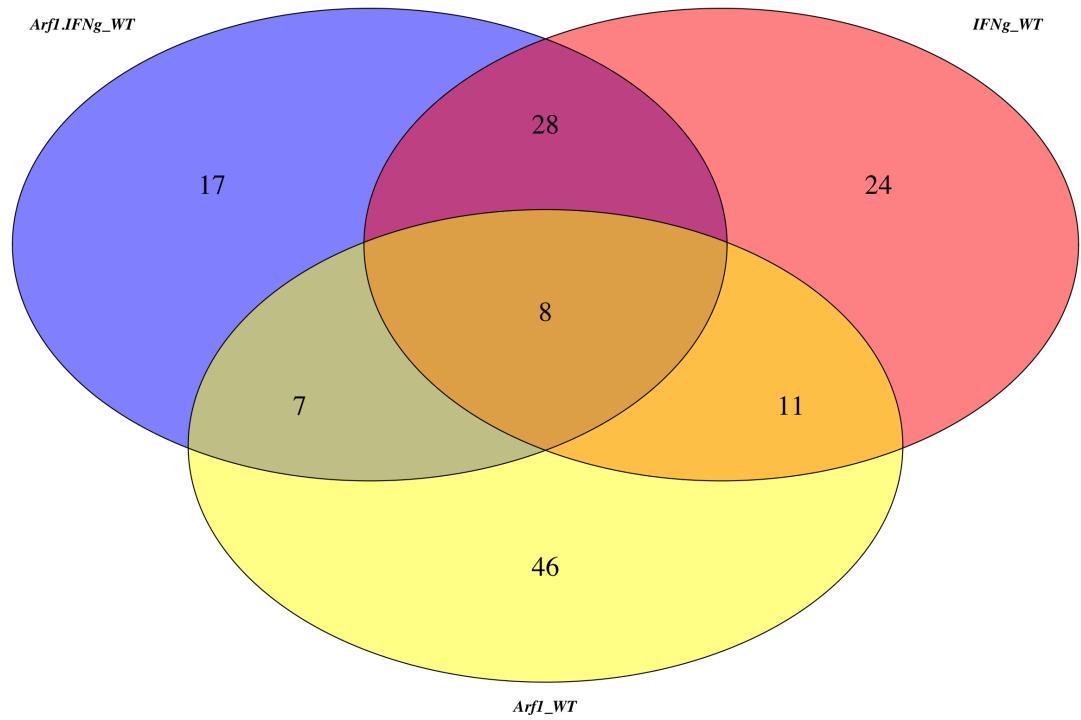
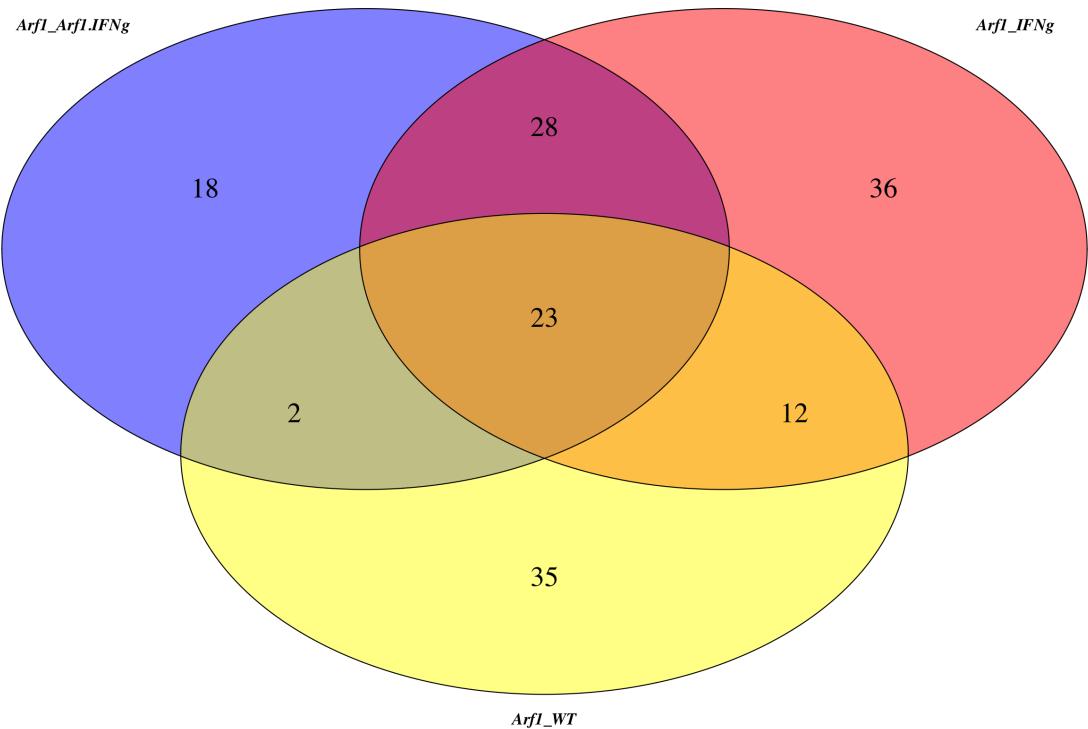




Cd68



## mouseRNAseq\_main microglia differential expression – DEGs common to all three contrasts



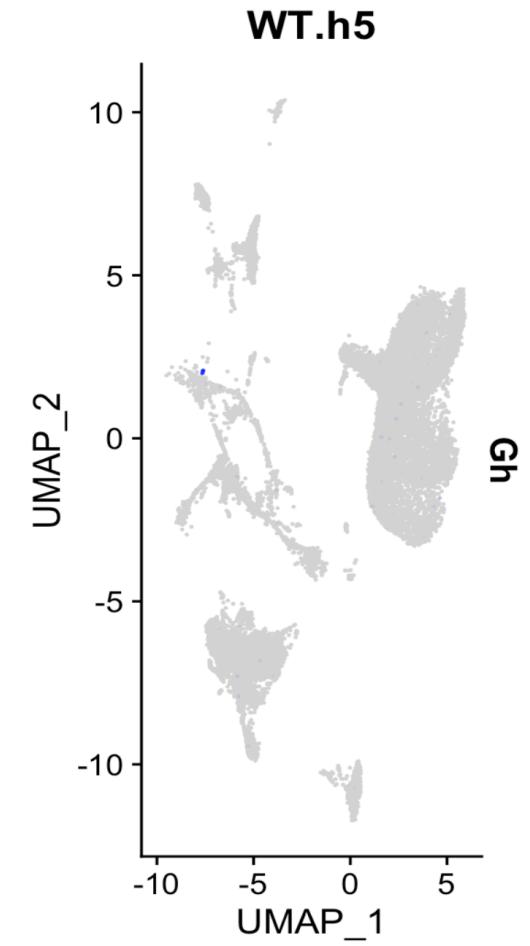
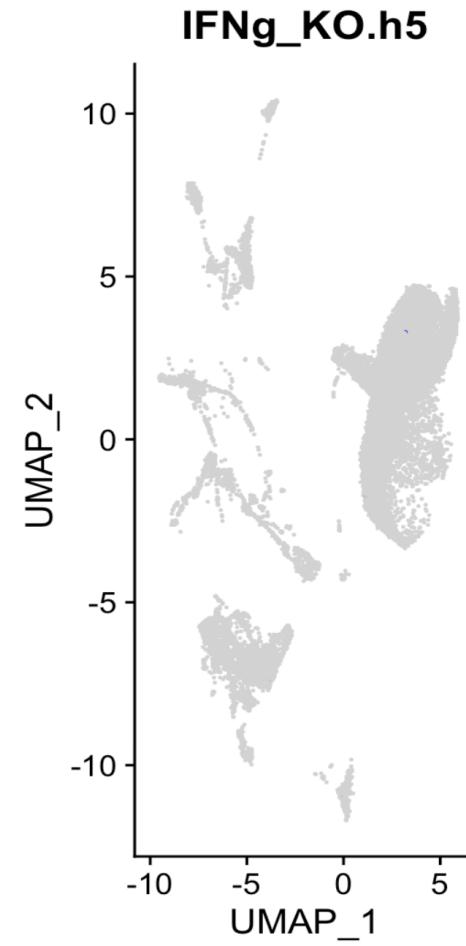
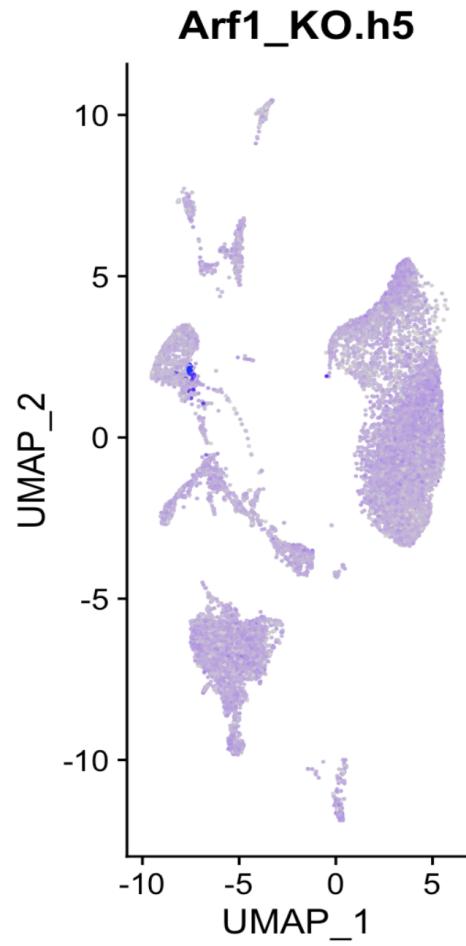
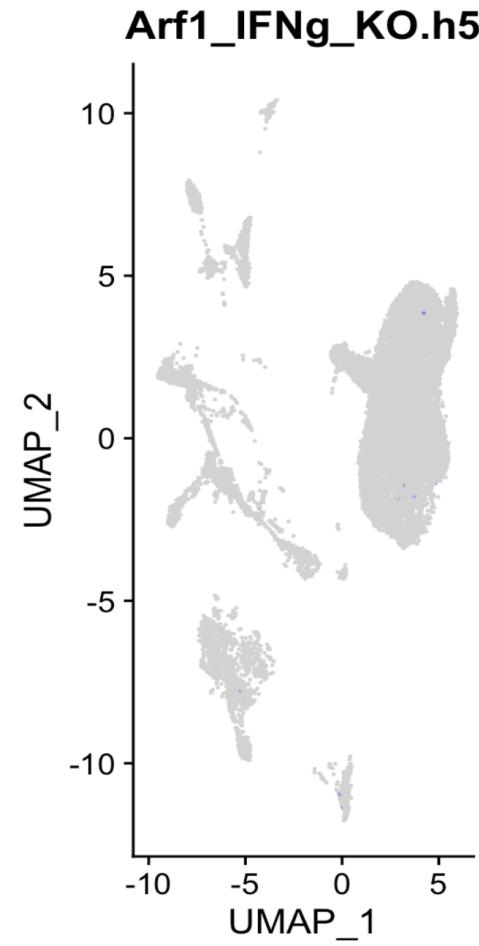
# mouseRNAseq\_main microglia differential expression – DEGs common to all three contrasts

	<b>Arf1 - Arf1.IFN<math>\gamma</math></b>	<b>Arf1 - IFN<math>\gamma</math></b>	<b>Arf1 - WT</b>
<b>Gh</b>	2.2601131	2.2580543	2.2539953
<b>Prl</b>	1.3881627	1.3878029	1.3561963
<b>Ttr</b>	0.7251019	0.6756903	0.9272252
<b>Apod</b>	0.6499153	0.5500515	0.5531819
<b>Pomc</b>	0.3739642	0.5384407	0.3896798
<b>Rbm3</b>	0.3182811	0.3837139	0.3750010
<b>Hbb-bs</b>	-0.2970969	0.3742991	-0.2946736
<b>Cx3cr1</b>	-0.3108693	-0.4587319	-0.3059742
<b>Gpr34</b>	-0.3946505	-0.4685851	-0.4441504
<b>Sparc</b>	-0.4760408	-0.5313699	-0.4724615
<b>Cst3</b>	-0.8293340	-0.5352697	-0.5062032
<b>Gm26825</b>	-0.9311943	-0.5626591	-0.2716750
<b>P2ry12</b>	-0.4047551	-0.6170945	0.2814050
<b>Marcks</b>	-0.2700342	-1.0073088	-0.3187753
<b>Ccl3</b>	0.5233011	-1.2355064	-0.2997791
<b>Siglech</b>	-0.2935748	-0.3230028	-0.2975228
<b>Tmem119</b>	-0.3321814	-0.2763223	0.3123555
<b>Lgals3bp</b>	0.3070131	0.2587596	0.2651458
<b>H2-D1</b>	0.2856554	-0.3015456	-0.2621052
<b>Ier3</b>	0.3555399	0.4715129	-0.2767758
<b>Cxcl2</b>	0.3010432	0.3712888	0.3039998
<b>S100a9</b>	-0.2961911	0.3466460	0.2864130
<b>Hspa1a</b>	-0.2897662	-0.2879384	0.2882253

**avg\_logFC**

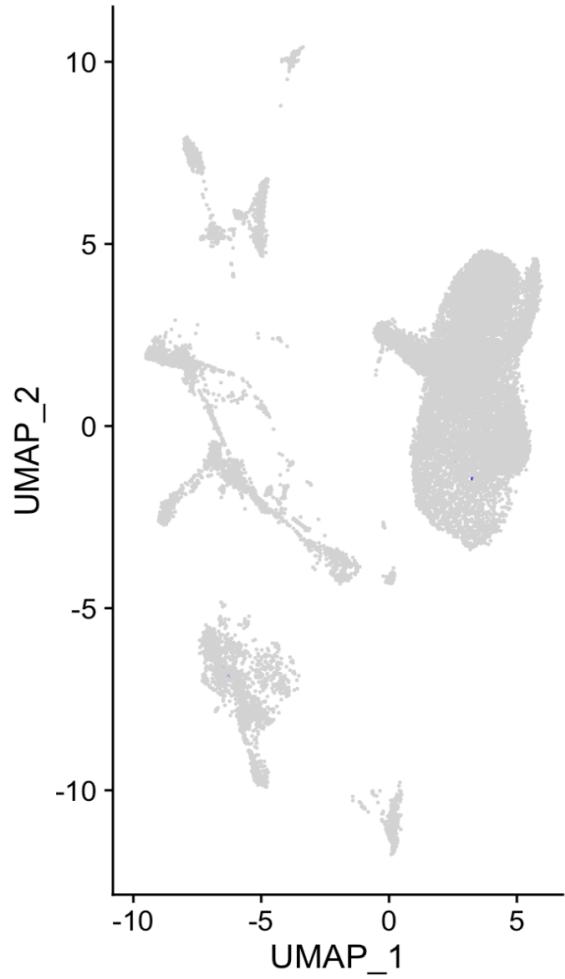
	<b>Arf1.IFN<math>\gamma</math> - WT</b>	<b>IFN<math>\gamma</math> - WT</b>	<b>Arf1 - WT</b>
<b>Gm26825</b>	1.1963401	1.5006522	0.9223484
<b>Apoe</b>	0.9011597	0.6284356	0.4004677
<b>Cst3</b>	0.3851837	0.5631584	-0.4441504
<b>Slc2a5</b>	-0.3746139	0.4015651	-0.5244208
<b>Fcrls</b>	-0.3986105	0.3451437	-0.2577561
<b>mt-Nd1</b>	0.2704647	0.3036750	0.3261354
<b>mt-Nd4</b>	0.2668789	-0.4349556	0.3015796
<b>mt-Co3</b>	0.2626265	-1.0109501	0.2651458

Gh

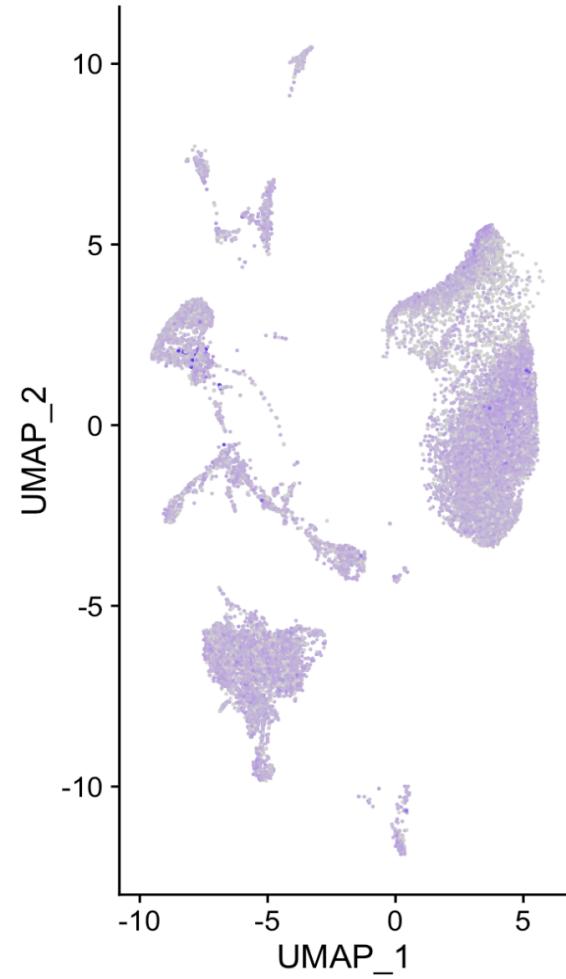


PrI

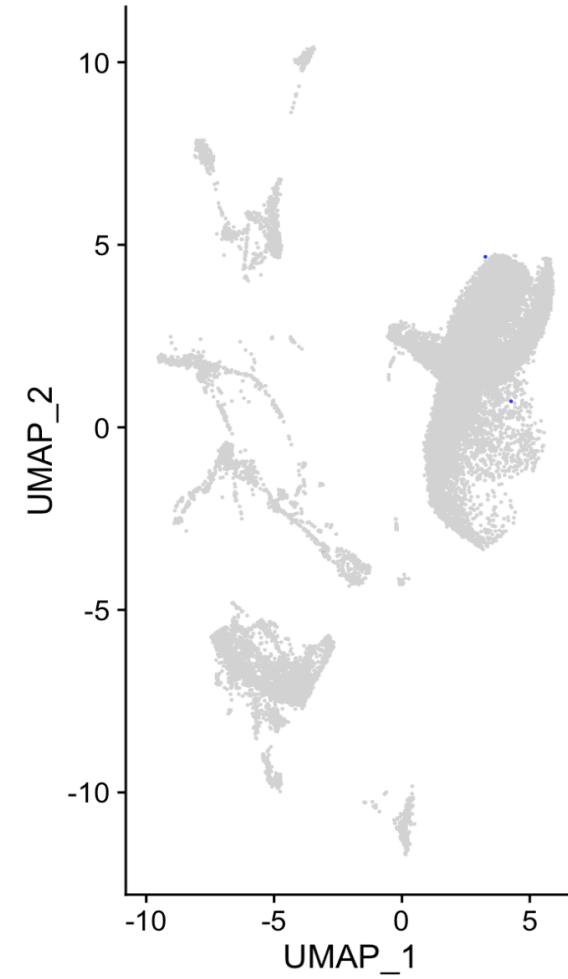
Arf1\_IFNg\_KO.h5



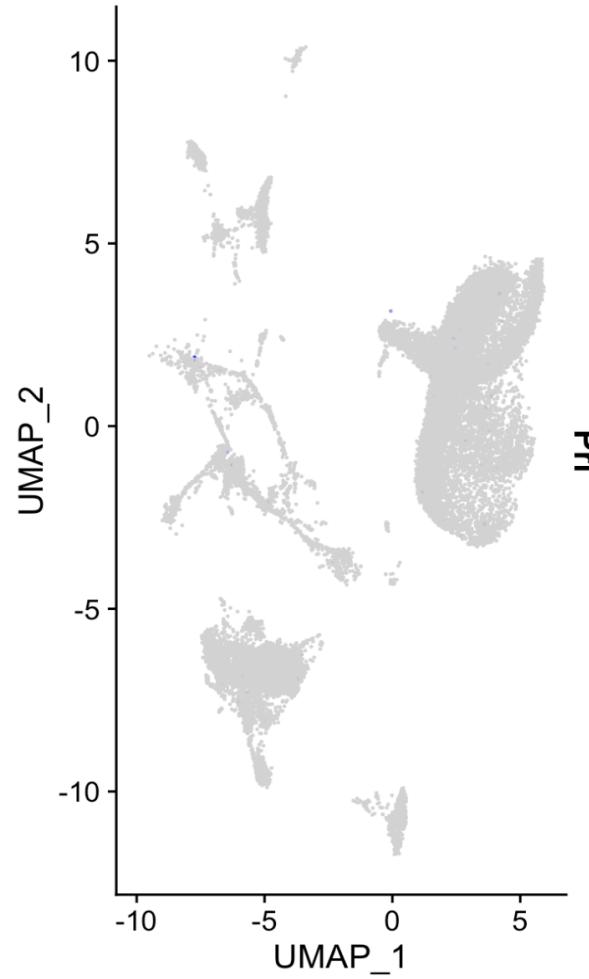
Arf1\_KO.h5



IFNg\_KO.h5

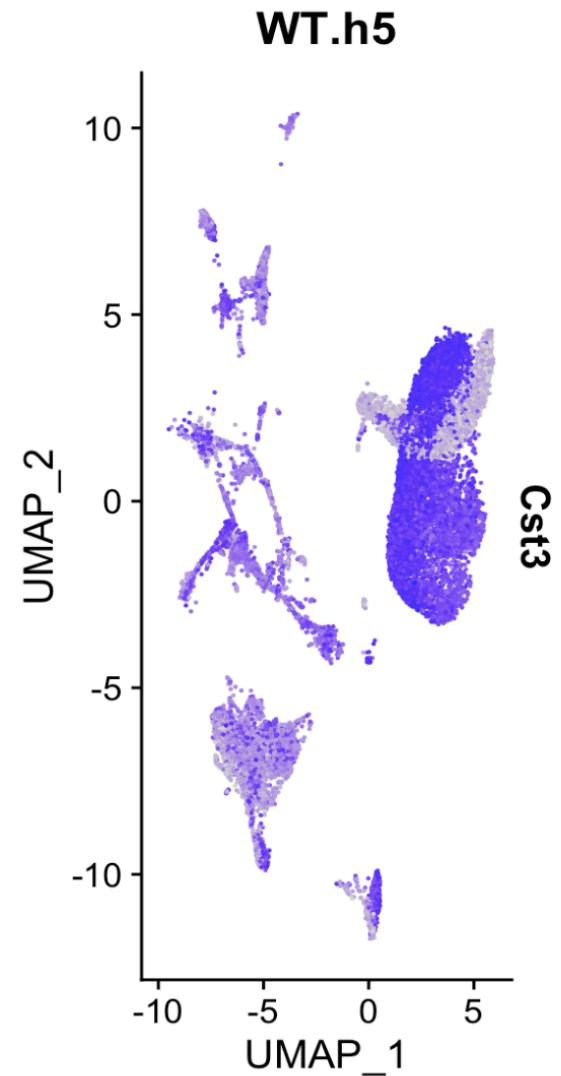
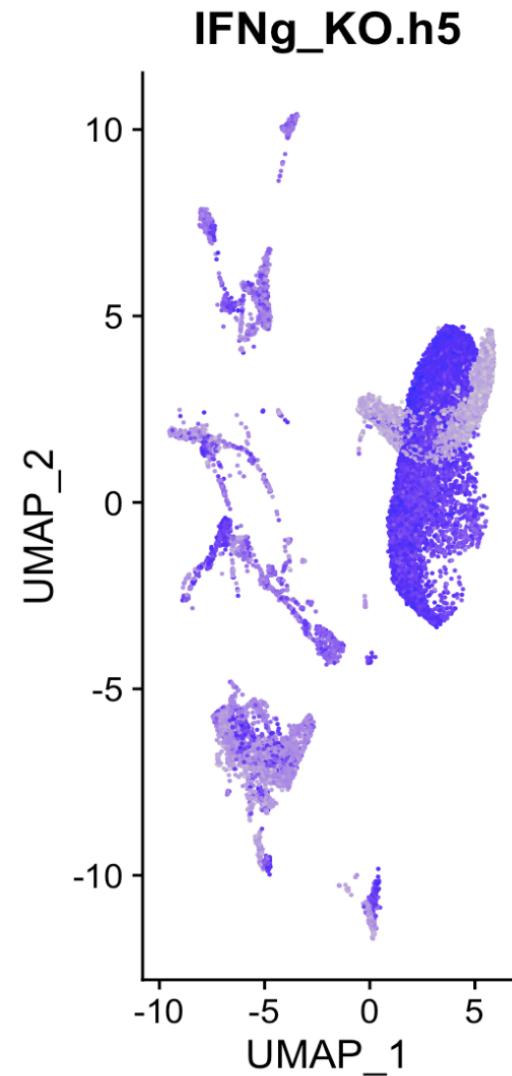
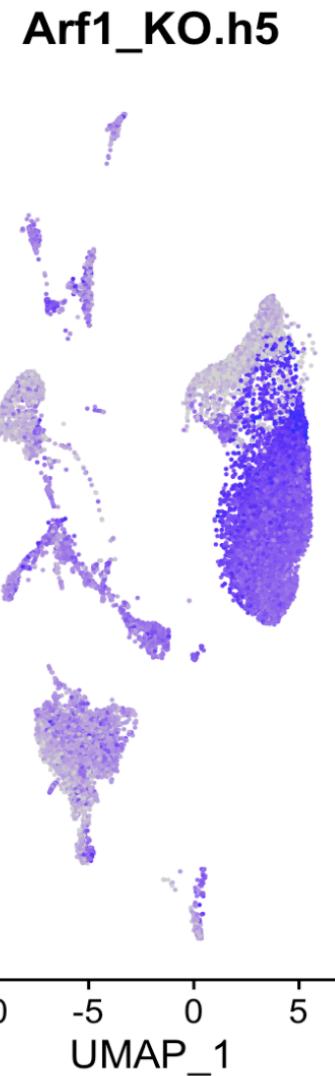
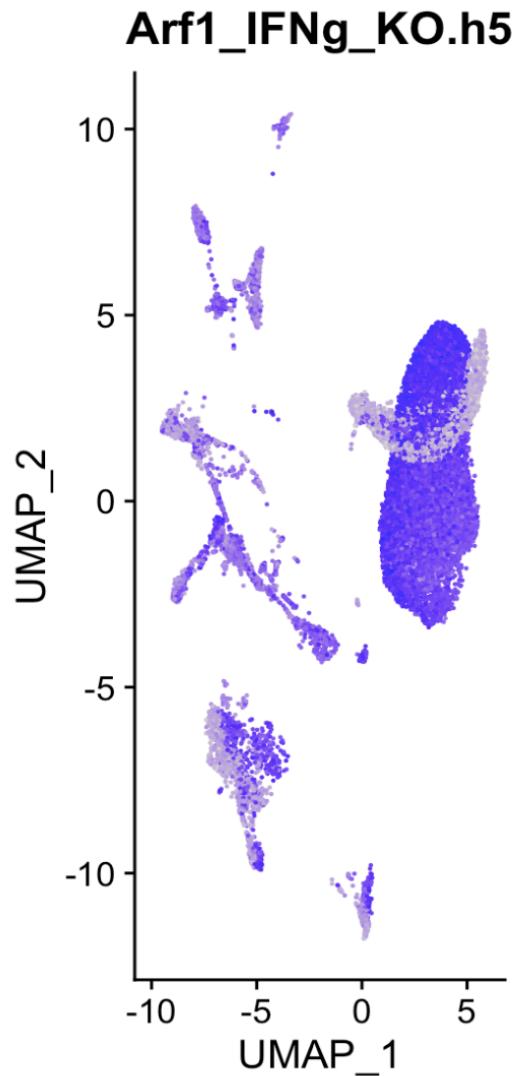


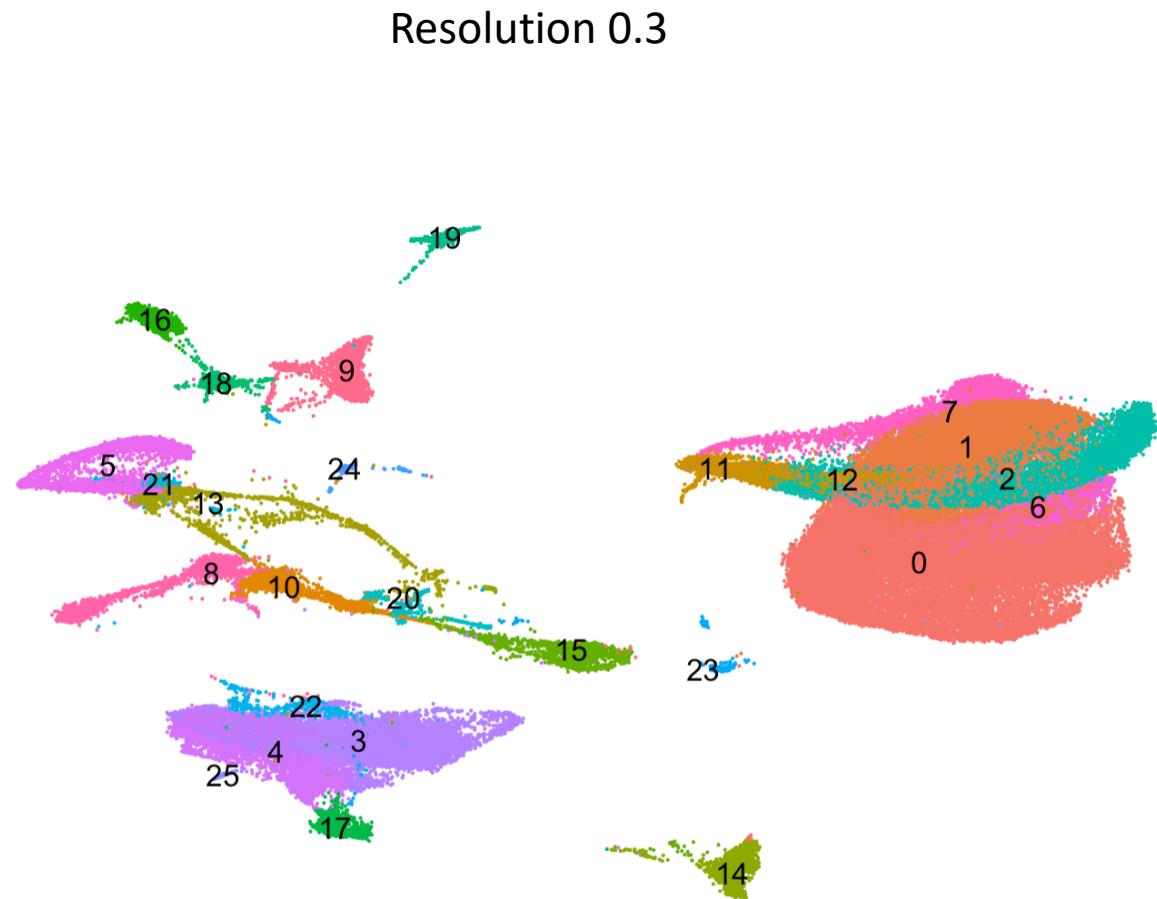
WT.h5



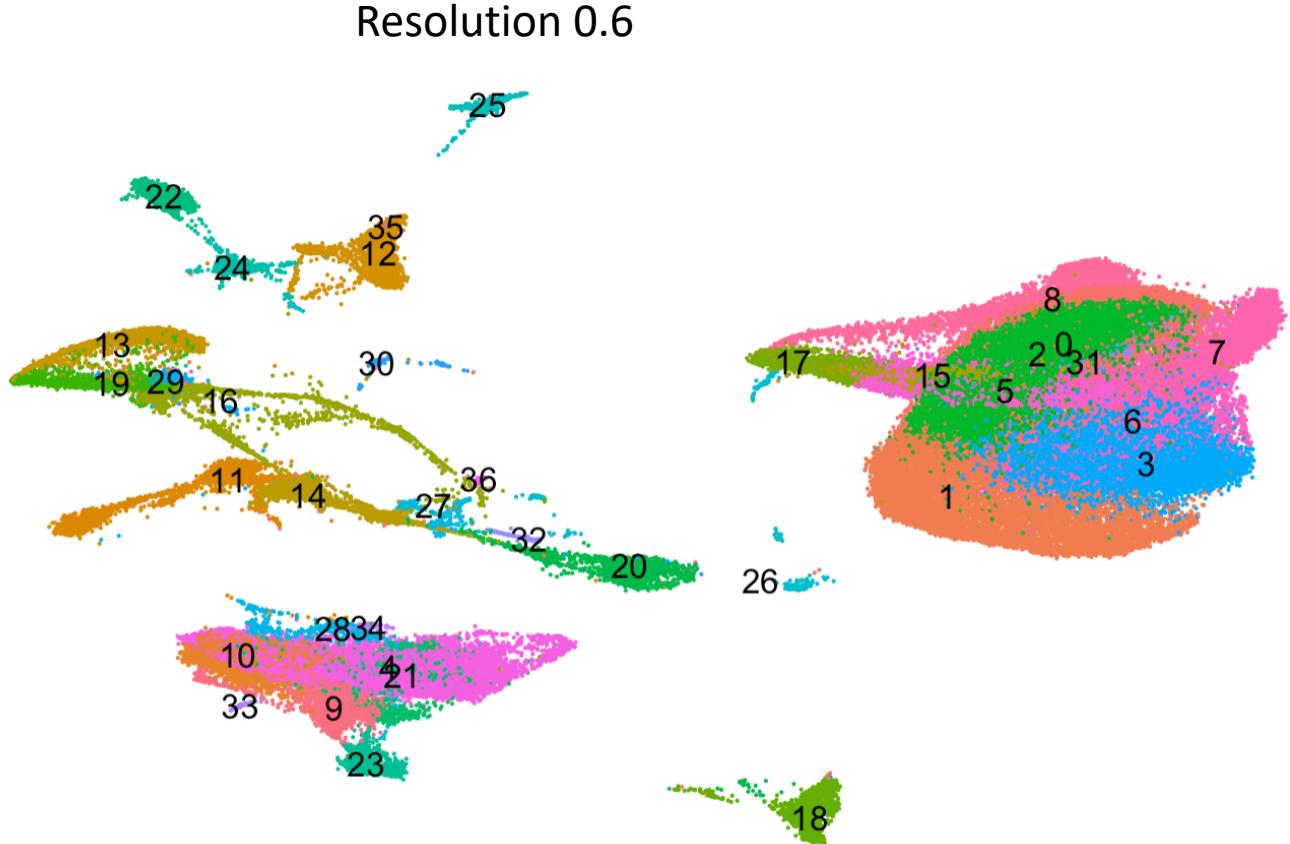
PrI

Cst3





mouseRNAseq_main	SCT_snn_res.0.3
Microglia	0
Microglia	1
NA	2
Microglia	3
NK cells	4
Oligodendrocytes	5
Microglia	6
Erythrocytes	7
Microglia	8
Epithelial cells	9
Monocytes	10
Erythrocytes	11
Microglia	12
Granulocytes	13
Microglia	14
Macrophages	15
Astrocytes	16
NK cells	17
Neurons	18
Epithelial cells	19
Fibroblasts	20
Neurons	21
Endothelial cells	22
Microglia	23
Fibroblasts	24
Microglia	25



mouseRNAseq_main	SCT_snn_res.0.6
Microglia	0
Microglia	1
Microglia	2
Microglia	3
NA	4
Erythrocytes	5
Microglia	6
NA	7
Erythrocytes	8
NK cells	9
T cells	10
Microglia	11
Epithelial cells	12
Oligodendrocytes	13
Monocytes	14
Microglia	15
Granulocytes	16
Erythrocytes	17
Microglia	18
Oligodendrocytes	19
Macrophages	20
Microglia	21
Astrocytes	22
NK cells	23
Neurons	24
Epithelial cells	25
Microglia	26
Fibroblasts	27
Endothelial cells	28
Neurons	29
Fibroblasts	30
Epithelial cells	31
Fibroblasts	32
Microglia	33
Endothelial cells	34
Epithelial cells	35
Monocytes	36

cell <fcctr>	Arf1_Freq <int>	Arf1_percentage <dbl>	WT_Arf1_Freq <int>	WT_Arf1_percentage <dbl>	IFNg_Freq <int>	IFNg_percentage <dbl>	Arf1_IFNg_Freq <int>	Arf1_IFNg_percentage <dbl>
aNSCs	401	0.02	248	0.01	267	0.01	203	0.01
B cells	294	0.01	335	0.01	171	0.01	240	0.01
Endothelial cells	2856	0.13	2715	0.11	2325	0.10	280	0.01
Ependymal	823	0.04	867	0.04	1198	0.05	815	0.03
Erythrocytes	1315	0.06	2069	0.08	2857	0.12	2547	0.11
Granulocytes	355	0.02	448	0.02	412	0.02	365	0.02
Macrophages	531	0.02	303	0.01	260	0.01	386	0.02
Microglia	8967	0.41	13493	0.55	12781	0.55	14001	0.60
Monocytes	656	0.03	998	0.04	705	0.03	844	0.04
NK cells	638	0.03	296	0.01	139	0.01	326	0.01
Oligodendrocytes	2459	0.11	160	0.01	229	0.01	744	0.03
qNSCs	248	0.01	487	0.02	284	0.01	526	0.02
T cells	1140	0.05	774	0.03	493	0.02	814	0.03

Total cells

Arf1\_IFNg\_KO.h5  
23484

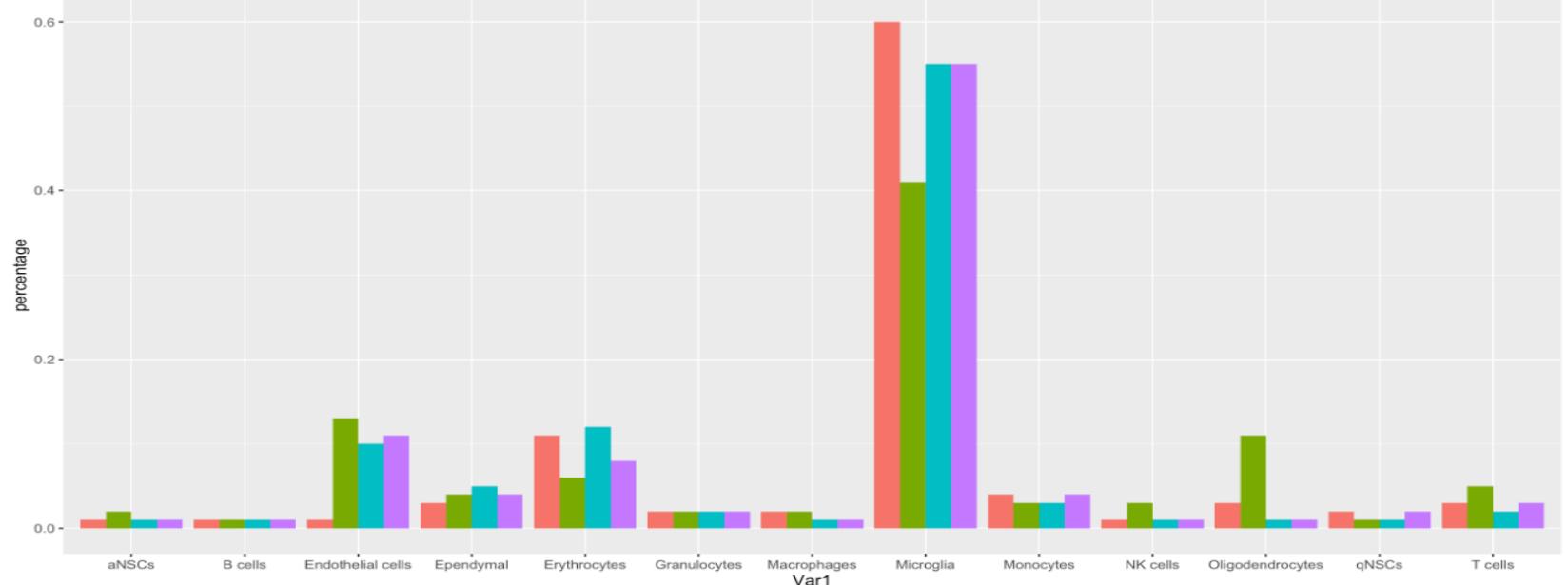
Arf1\_KO.h5  
22036

IFNg\_KO.h5  
23166

WT.h5  
24531

Arf1\_IFNg\_KO.h5 Arf1\_KO.h5 IFNg\_KO.h5 WT.h5

sample Arf1\_IFNg\_KO\_cells Arf1\_KO\_cells IFNg\_KO\_cells WT\_cells



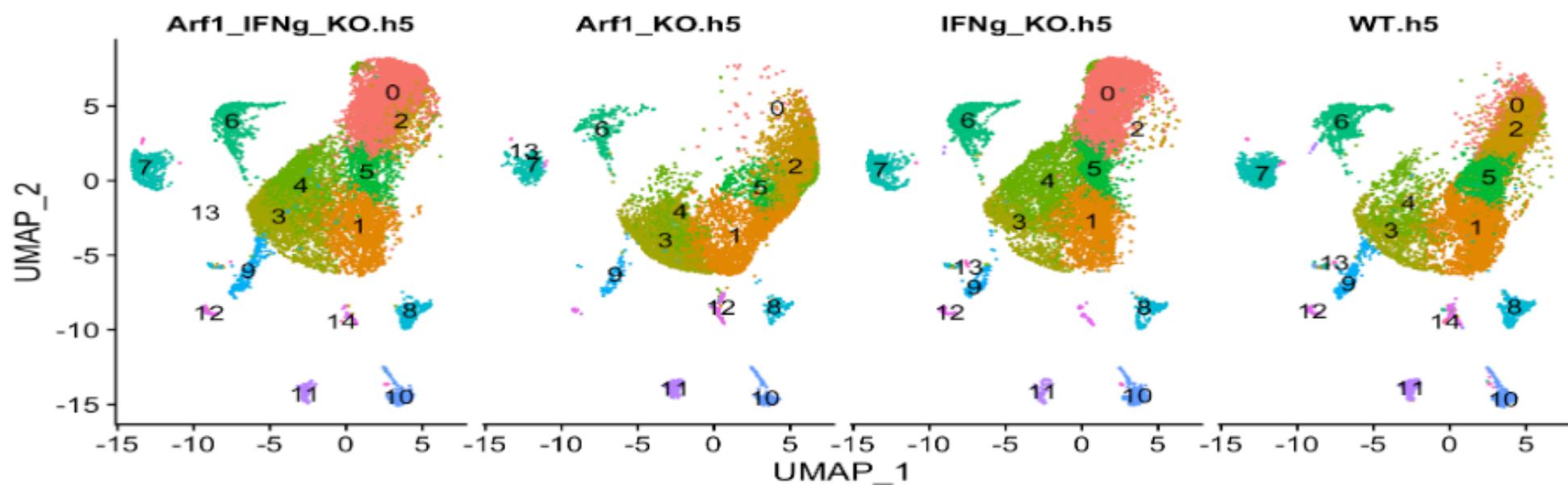
	Arf1_IFNg_KO.h5	Arf1_KO.h5	IFNg_KO.h5	WT.h5
Microglia	13733	8707	12802	13420
Oligodendrocytes	745	2453	233	170
Erythrocytes	2814	1641	3255	2358
Astrocytes	552	242	285	477
Epithelial cells	906	984	1348	950
NK cells	322	611	145	325
T cells	852	1197	507	796
Macrophages	1026	1124	524	665
Granulocytes	358	349	339	439
Fibroblasts	450	227	182	164
Endothelial cells	249	2837	2320	2721
Neurons	21	123	17	63
Monocytes	644	685	552	1002
B cells	239	293	173	341
Hepatocytes	30	39	29	19
Cardiomyocytes	20	8	1	9
Dendritic cells	7	10	8	14
Adipocytes	2	6	3	0

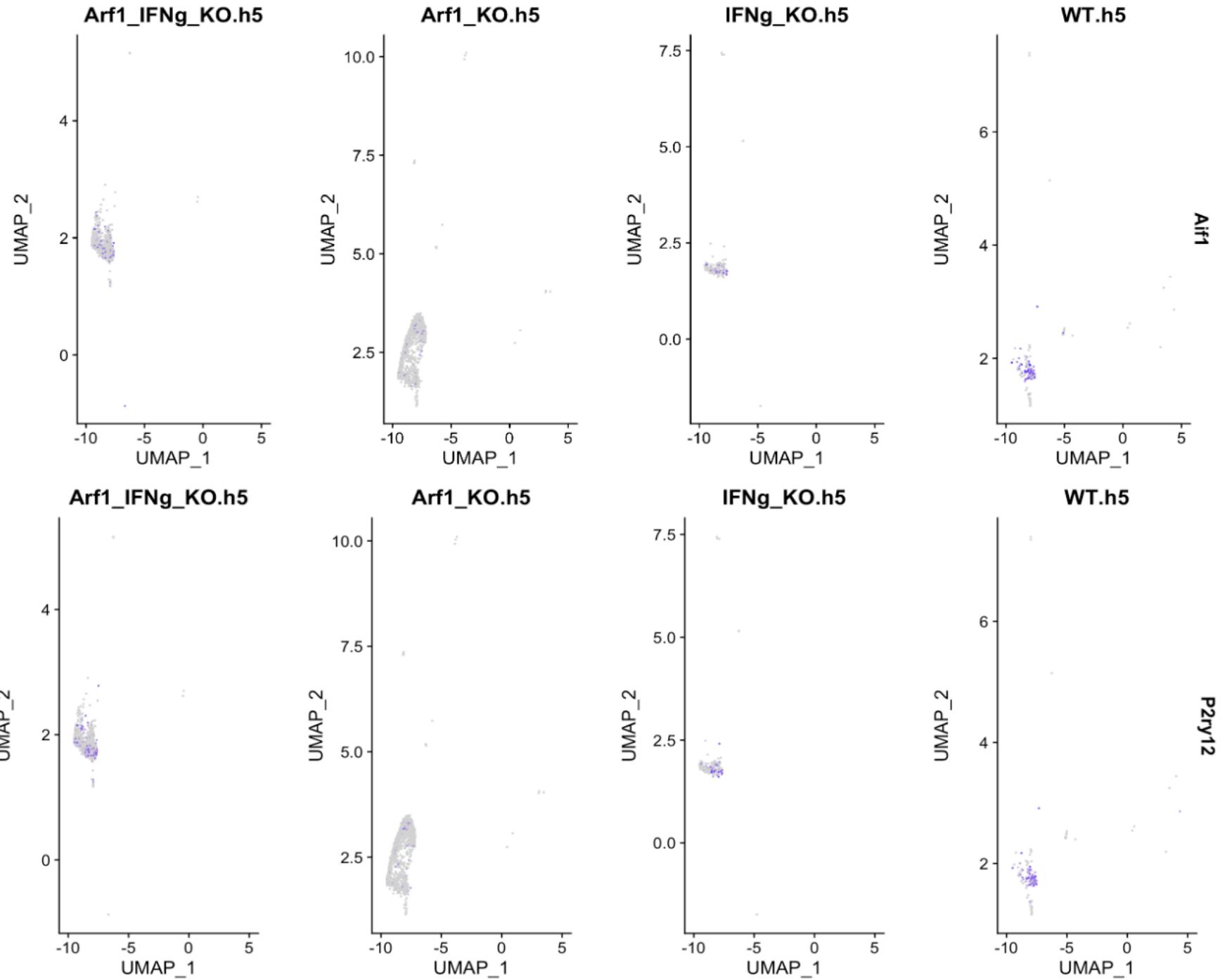
Select only the microglial cells annotated by  
mouseRNAseq\_main, reclusterd the cells

	Arf1_IFNg_KO.h5	Arf1_KO.h5	IFNg_KO.h5	WT.h5
Microglia	13733	8707	12802	13420

	Arf1_IFNg_KO.h5	Arf1_KO.h5	IFNg_KO.h5	WT.h5
0	5155	52	4913	313
1	1540	1904	2054	2868
2	408	2454	69	3870
3	1986	1796	884	804
4	1810	938	1626	760
5	710	329	1227	1789
6	559	258	772	767
7	466	303	361	699
8	333	137	256	387
9	331	78	165	386
10	183	131	295	309
11	134	168	85	239
12	105	153	74	125
13	12	6	21	42
14	1	0	0	62

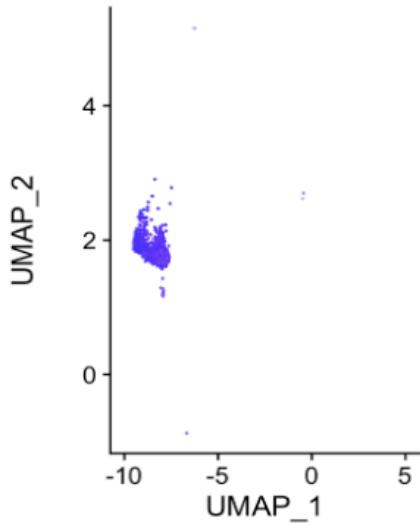
Number of  
microglia cells  
in each cluster  
split by sample



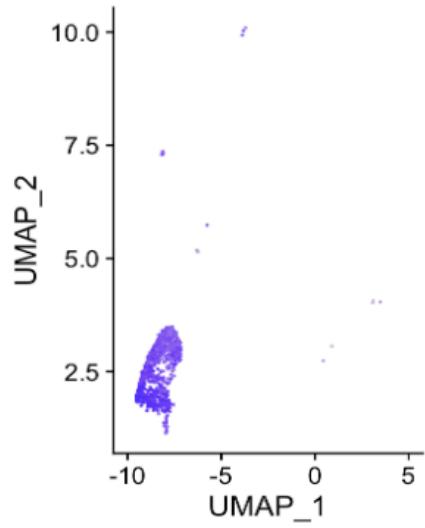


Checking expression of microglia markers in the cells labelled as Oligodendrocytes – weak but nonzero expression

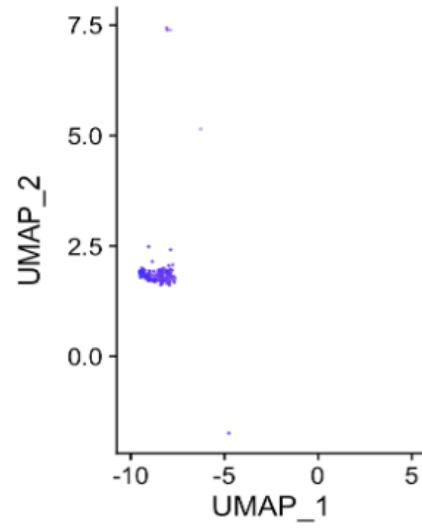
**Arf1\_IFNg\_KO.h5**



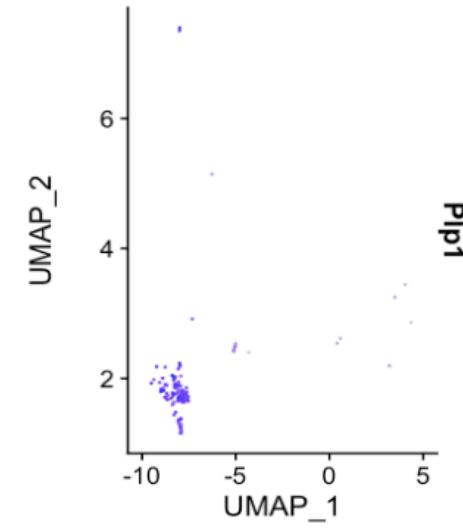
**Arf1\_KO.h5**



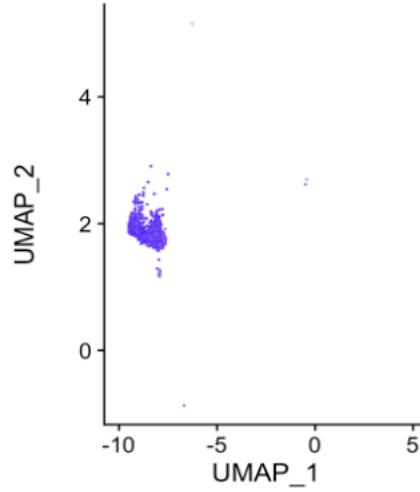
**IFNg\_KO.h5**



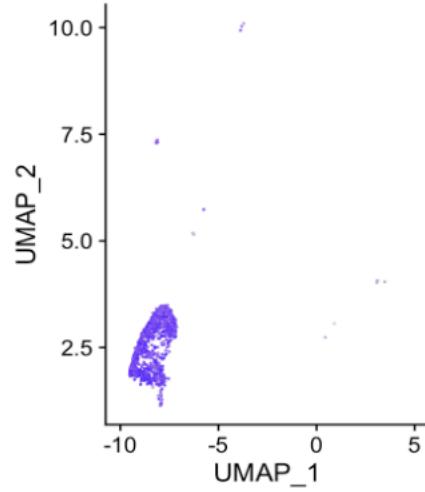
**WT.h5**



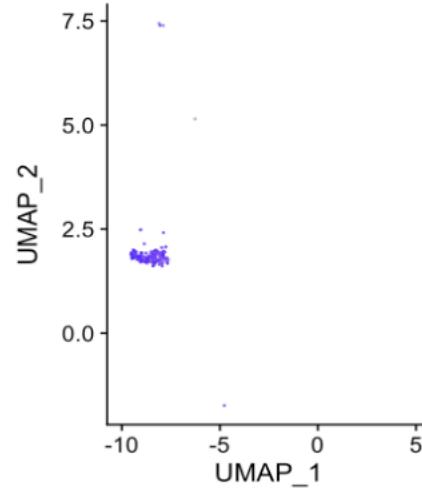
**Arf1\_IFNg\_KO.h5**



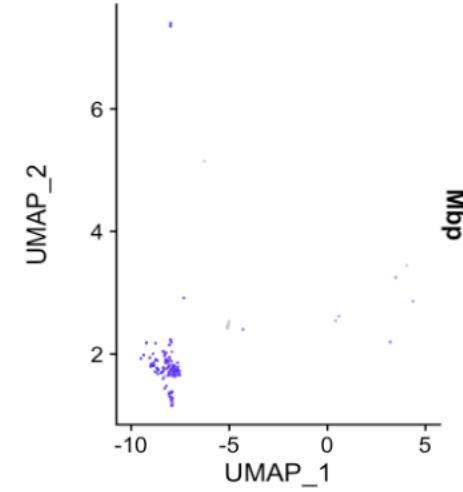
**Arf1\_KO.h5**



**IFNg\_KO.h5**

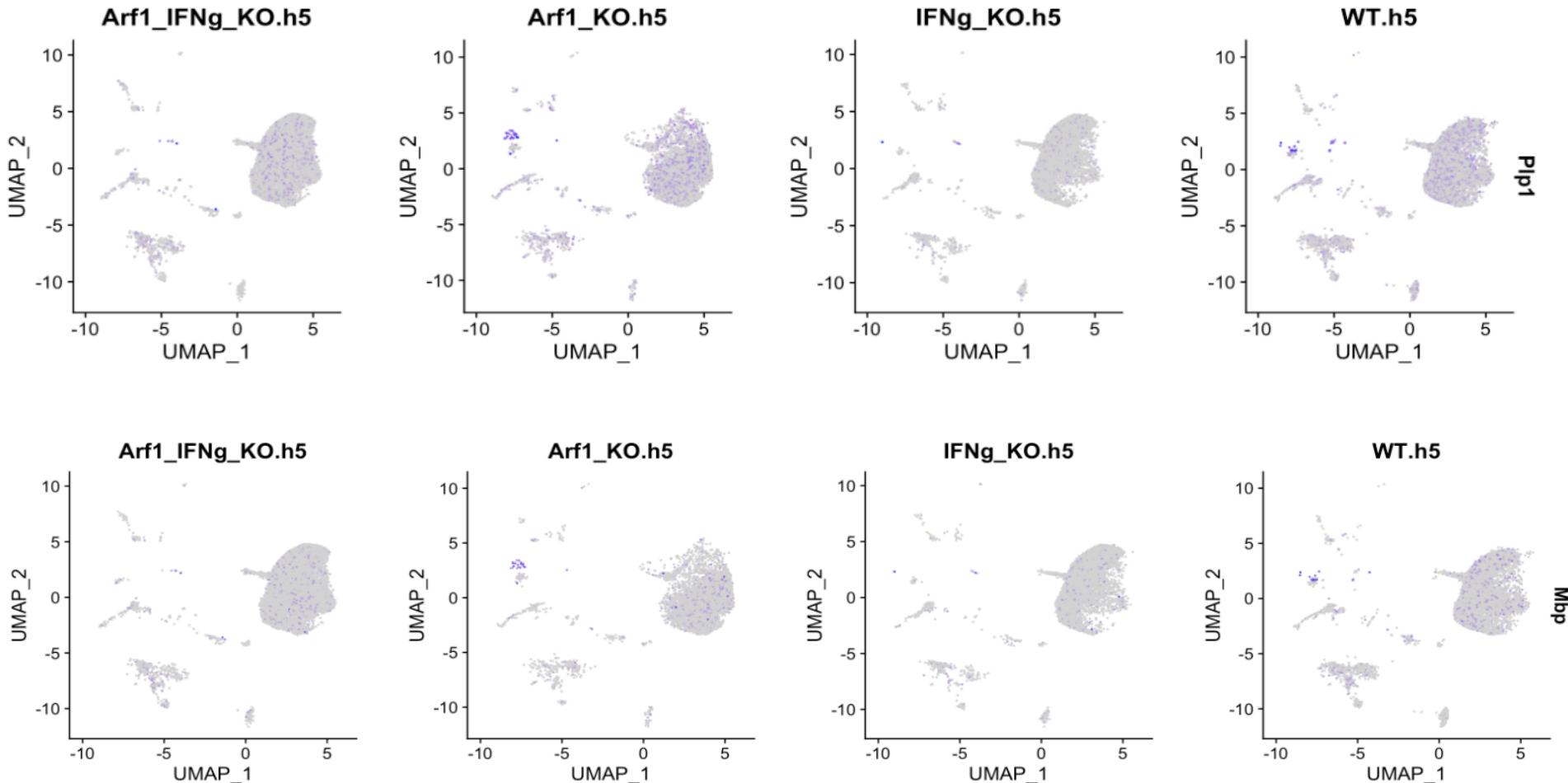


**WT.h5**

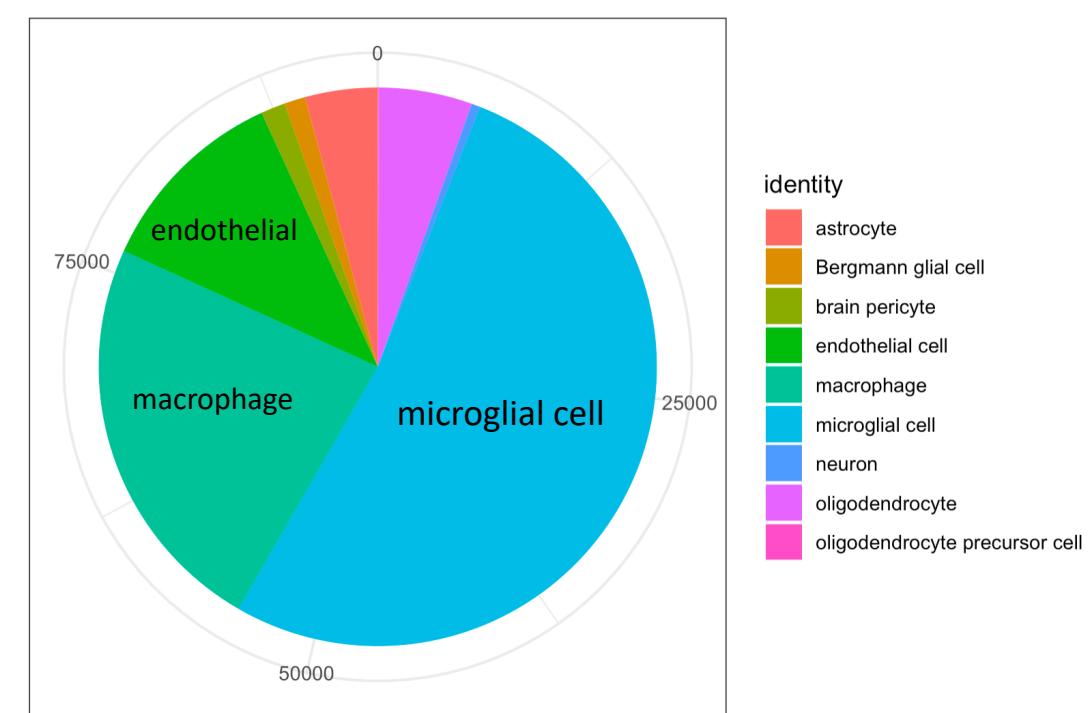
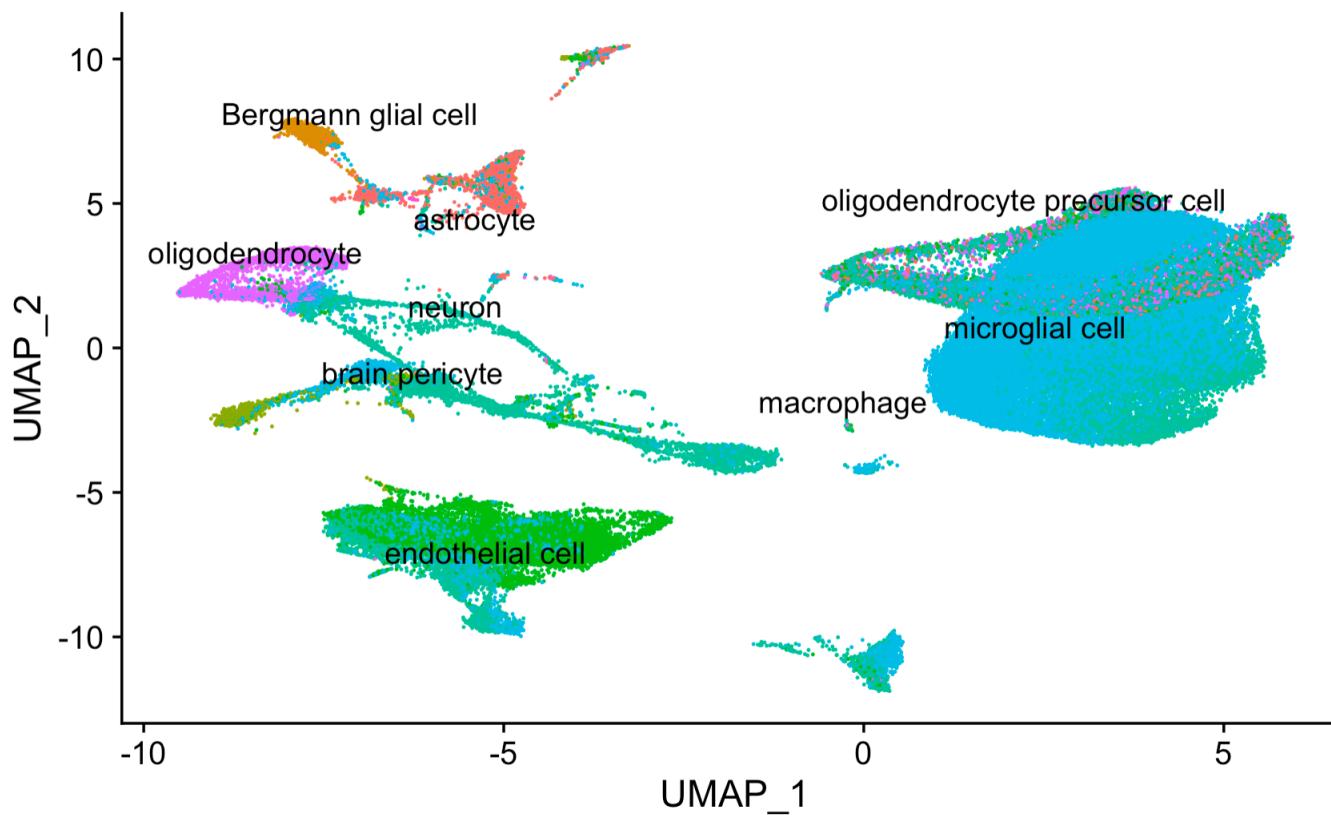


Checking expression of Oligodendrocytes markers in Oligodendrocytes – high expression as expected

## Checking expression of Oligodendrocytes markers in Microglia – weak but nonzero expression

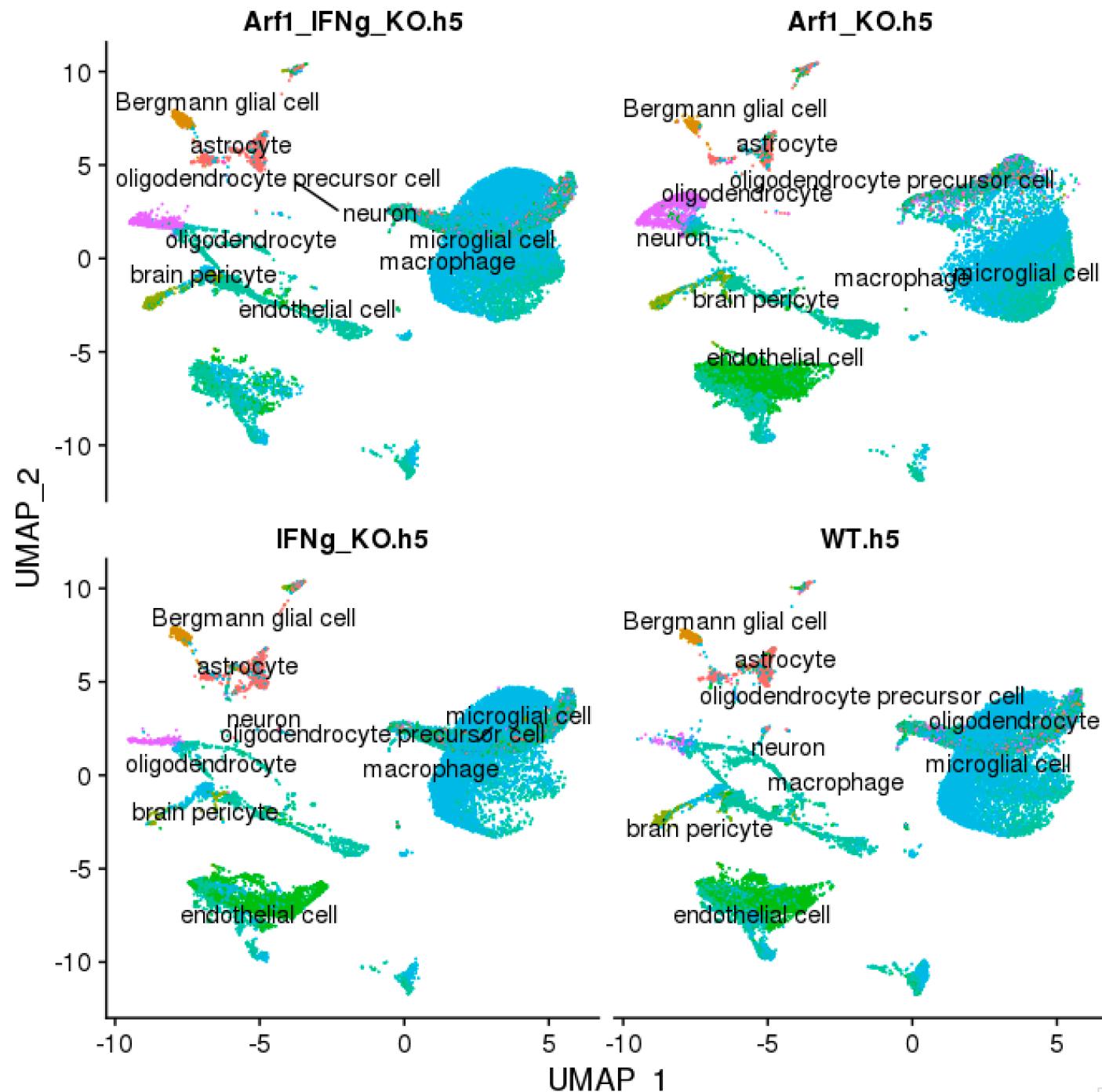


# Cell identity annotation by Tabula Muris – over half of the cells are microglia

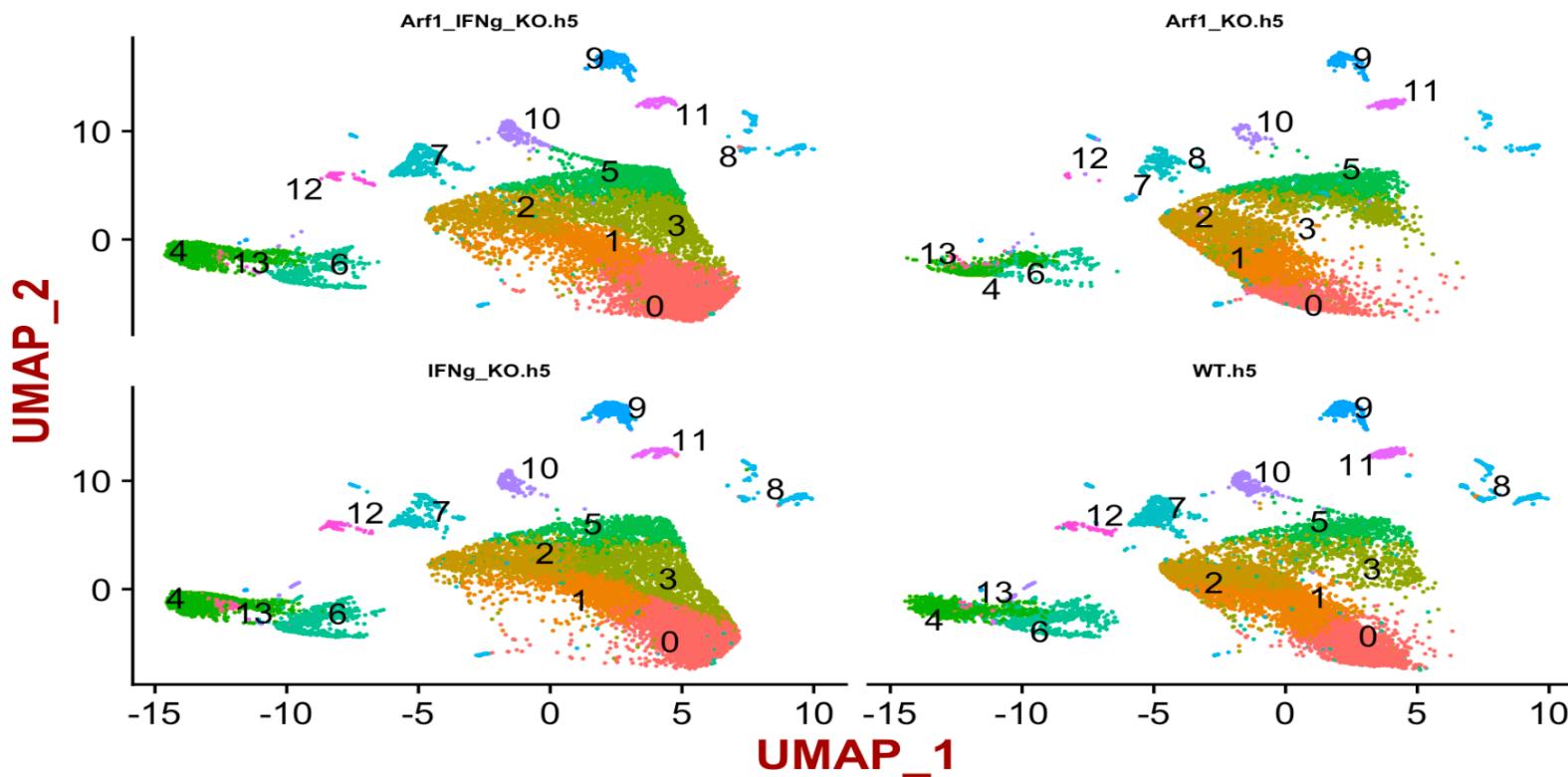


## Cell identity and count by sample

cell identity	Arf1_IFNg_KO.h5	Arf1_KO.h5	IFNg_KO.h5	WT.h5
astrocyte	1136	554	1177	1049
Bergmann glial cell	380	230	230	297
brain pericyte	461	349	241	266
endothelial cell	788	3473	3053	3276
macrophage	6387	6463	3853	5193
microglial cell	13238	7683	14105	13763
neuron	56	282	58	81
oligodendrocyte	1023	2988	438	602
oligodendrocyte precursor cell	15	14	11	4
Total	23484	22036	23166	24531

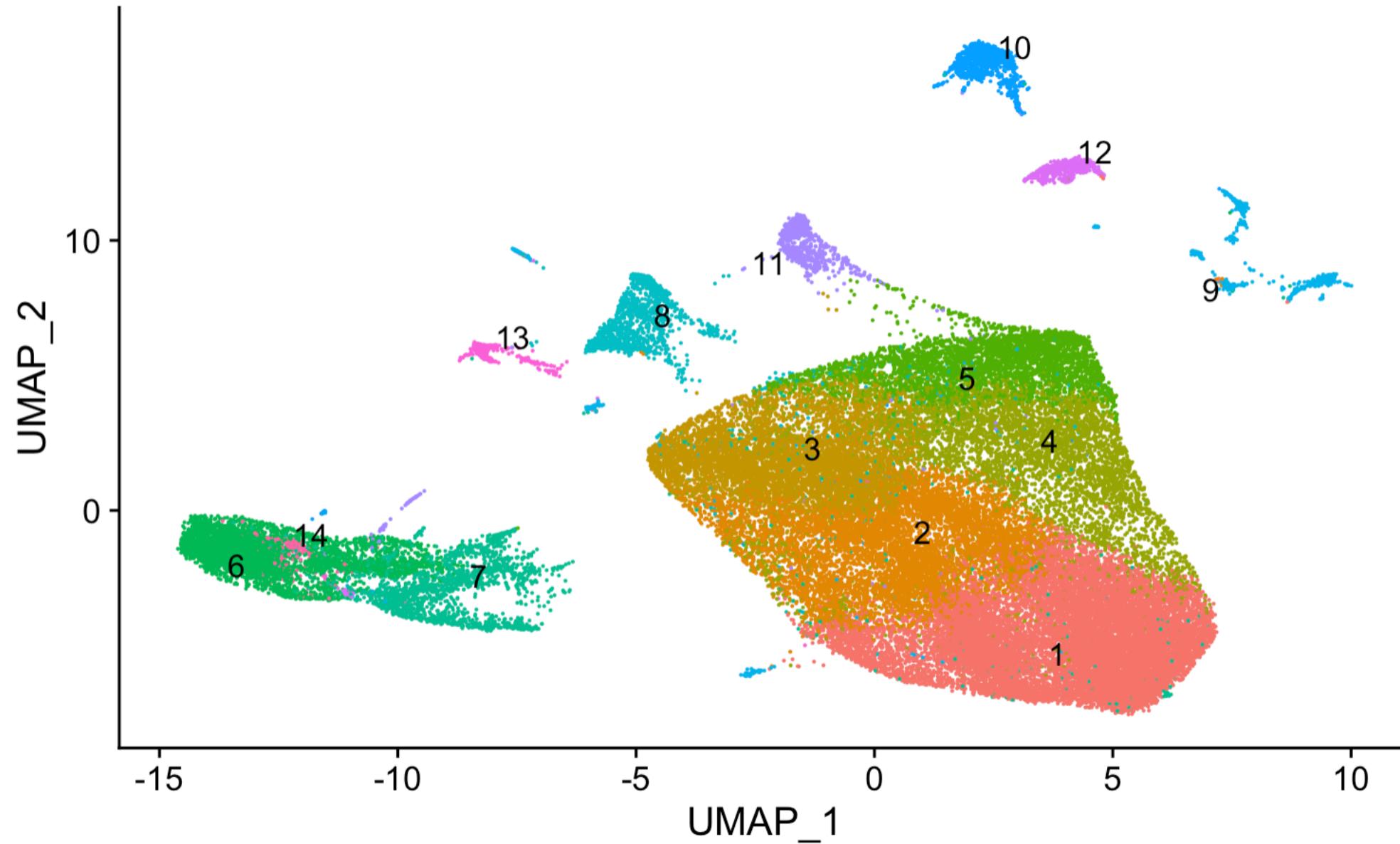


# Only microglia cells – reshaped with 0.3 resolution into 14 microglia subclusters



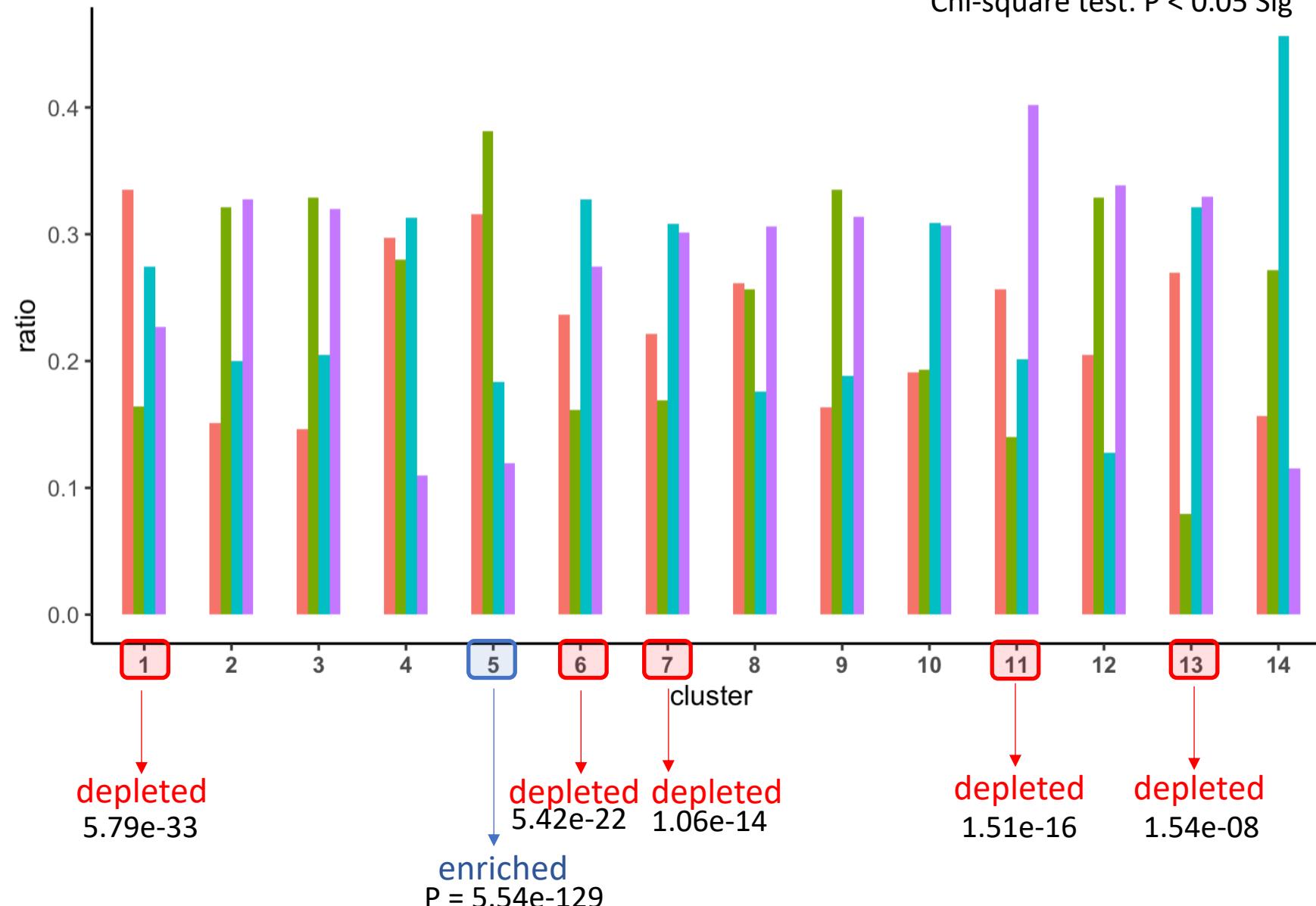
Number of microglial cells for each cluster in each sample

microglia_cluster	Arf1_IFNg_KO.h5	Arf1_KO.h5	IFNg_KO.h5	WT.h5
0	5081	1442	4429	3577
1	1221	1507	1724	2754
2	1030	1345	1537	2347
3	1698	930	1908	654
4	995	393	1469	1201
5	1428	1000	883	560
6	502	223	745	710
7	387	220	276	471
8	206	246	254	413
9	209	123	361	350
10	210	67	176	343
11	136	127	90	234
12	90	15	114	114
13	45	45	139	35
total	13238	7683	14105	13763



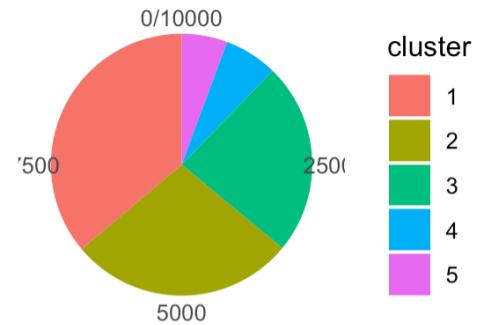
Sample Arf1\_IFNg\_KO Arf1\_KO IFNg\_KO WT

Chi-square test: P < 0.05 Sig

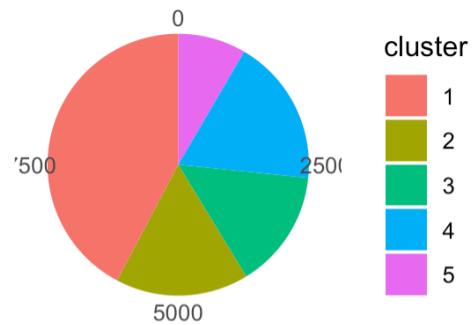


## Relative proportion cluster

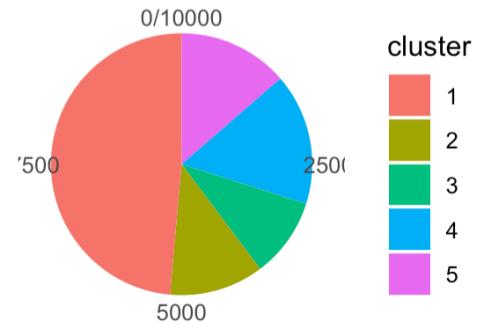
WT



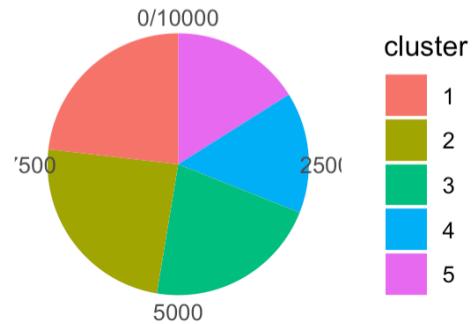
IFNg\_KO



Arf1\_IFNg\_KO



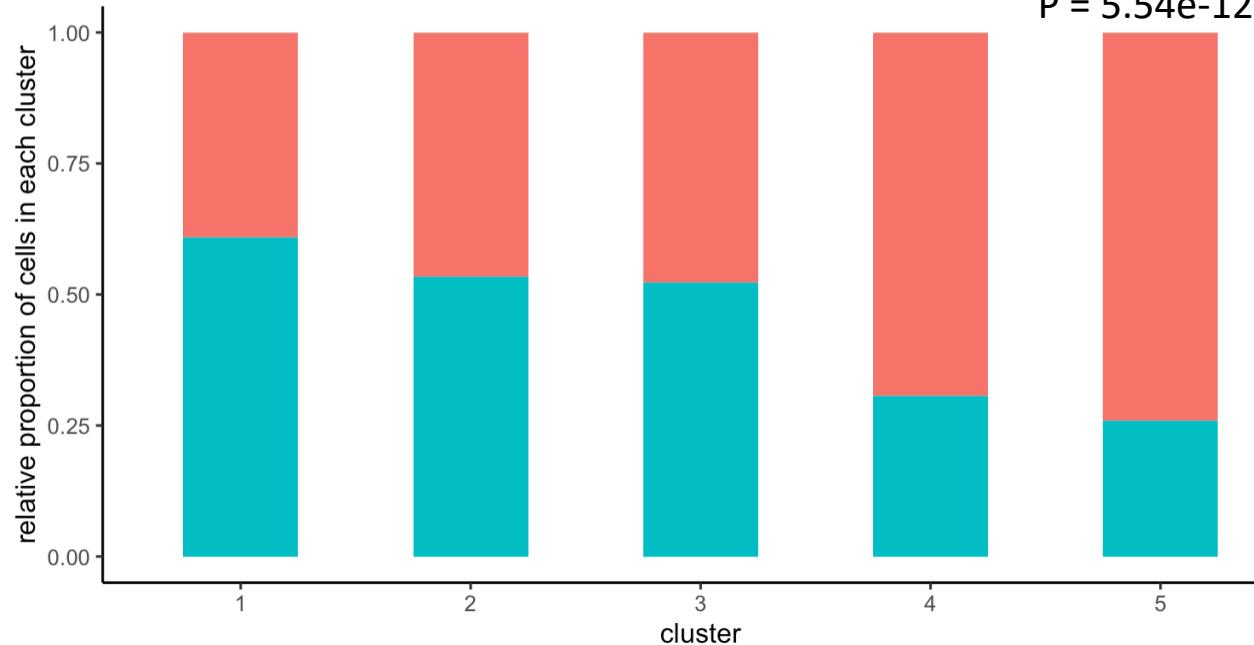
Arf1\_KO

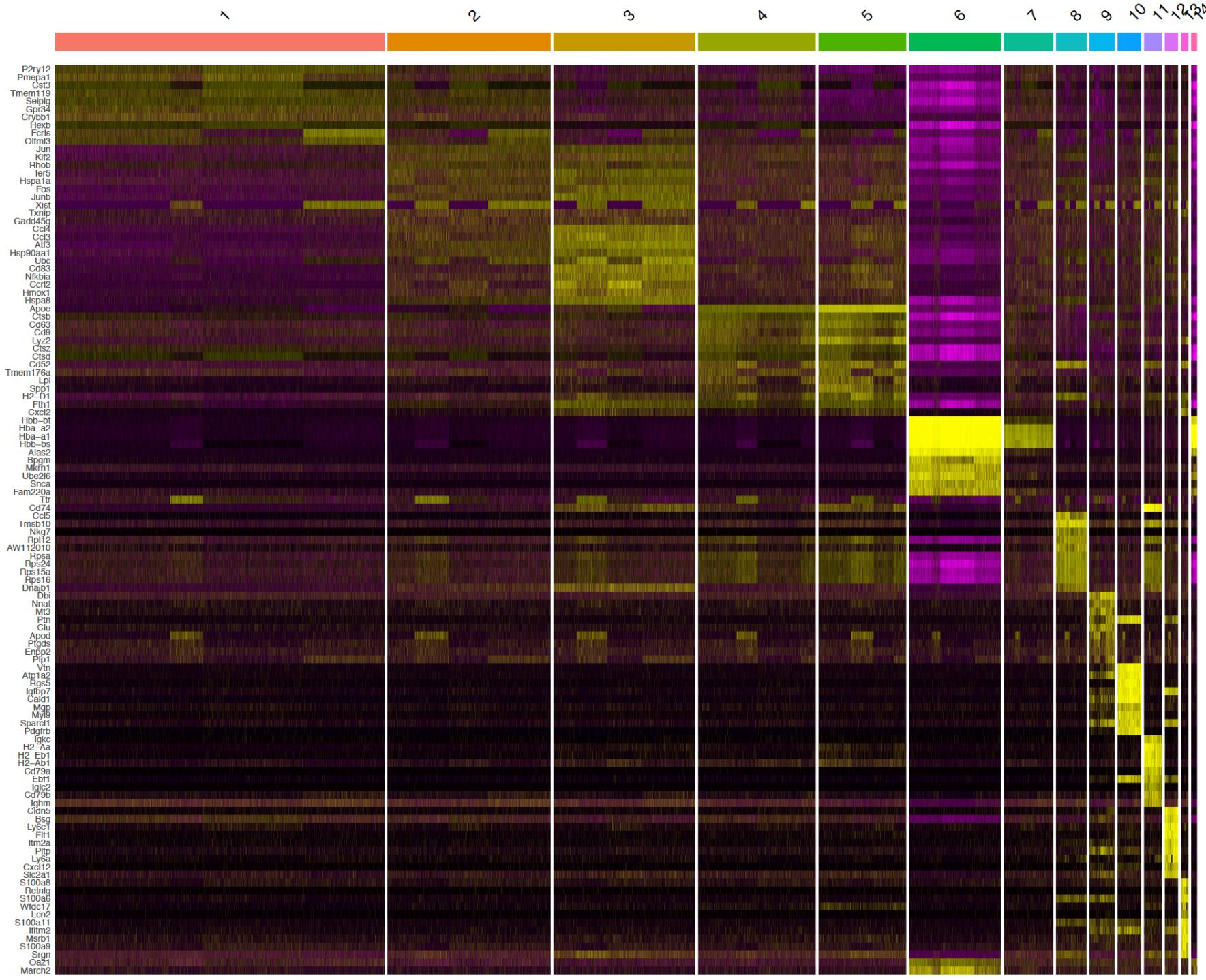


significantly higher proportion of cells in Arf1-KO cluster5

SampleName Arf1\_KO WT

$P = 5.54e-129$



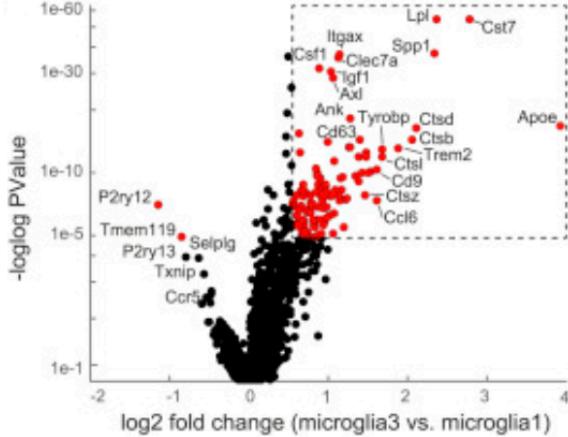


Markers (top 10 by average fold change)  
for each microglia  
subcluster

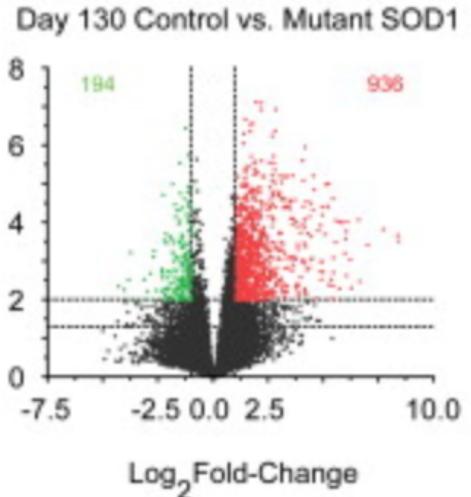
Article

# A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease

Hadas Keren-Shaul <sup>1, 6</sup>, Amit Spinrad <sup>1, 2, 6</sup>, Assaf Weiner <sup>1, 3, 6</sup>✉, Orit Matcovitch-Natan <sup>1, 2, 6</sup>, Raz Dvir-Szternfeld <sup>2</sup>, Tyler K. Ulland <sup>4</sup>, Eyal David <sup>1</sup>, Kuti Baruch <sup>2</sup>, David Lara-Astaiso <sup>1</sup>, Beata Toth <sup>5</sup>, Shalev Itzkovitz <sup>5</sup>, Marco Colonna <sup>4</sup>, Michal Schwartz <sup>2, 7</sup>✉, Ido Amit <sup>1, 7, 8</sup>✉

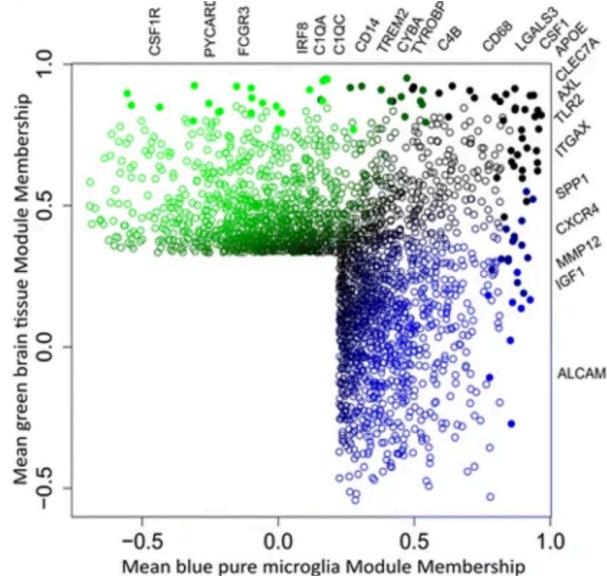


A Neurodegeneration-Specific Gene-Expression Signature of Acutely Isolated Microglia from an Amyotrophic Lateral Sclerosis Mouse Model  
Isaac M. Chiu • Emiko T.A. Morimoto • Hani Goodarzi • ... Saeed Tavazoie • Richard M. Myers • Tom Maniatis ✉ • Show all authors

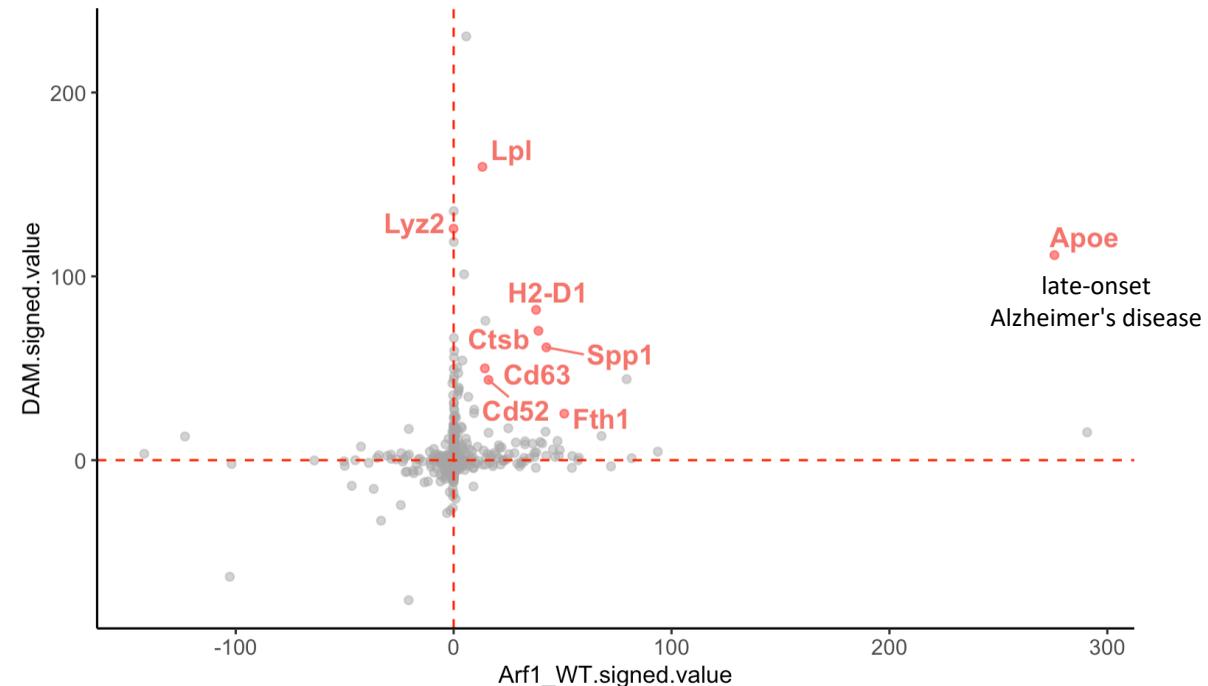
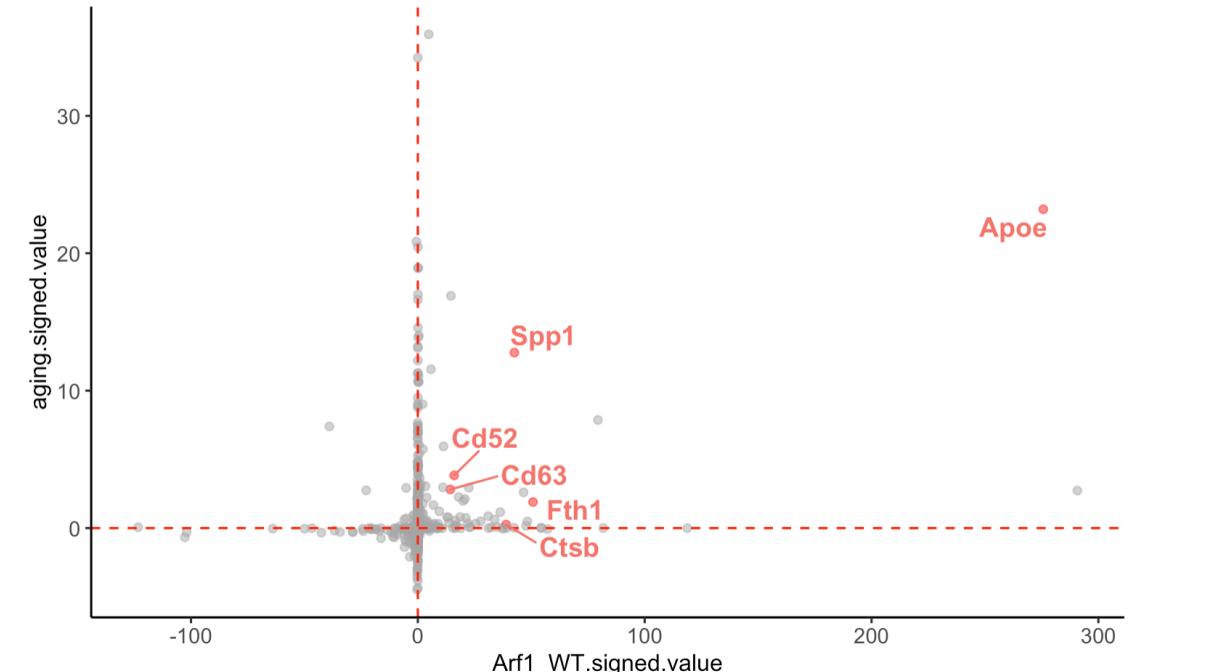
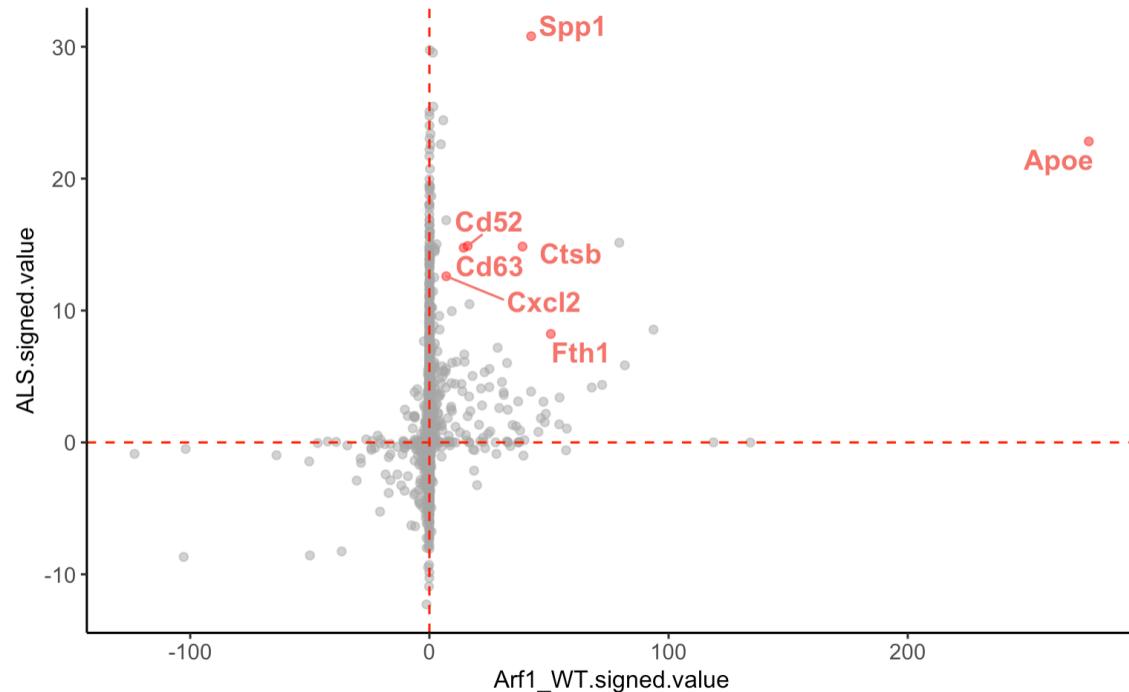


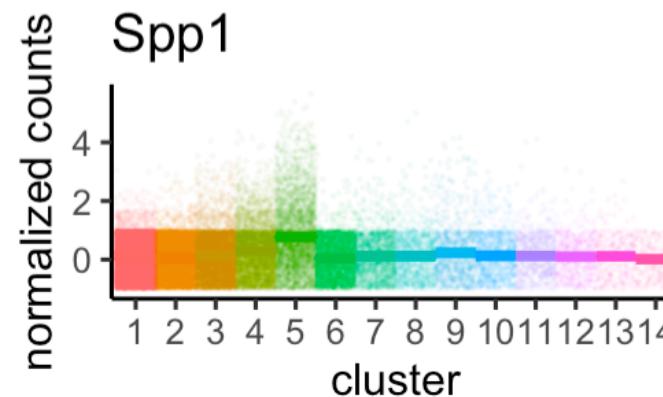
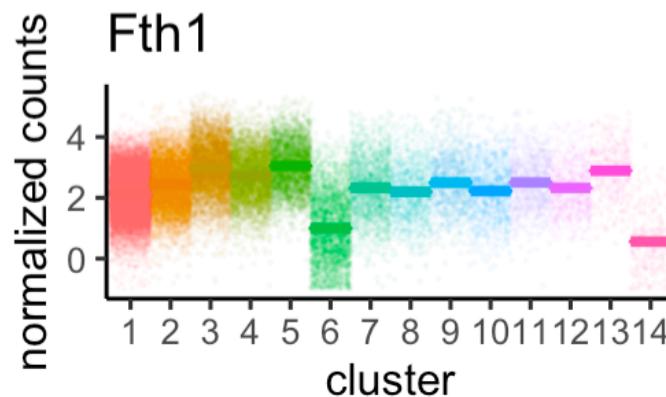
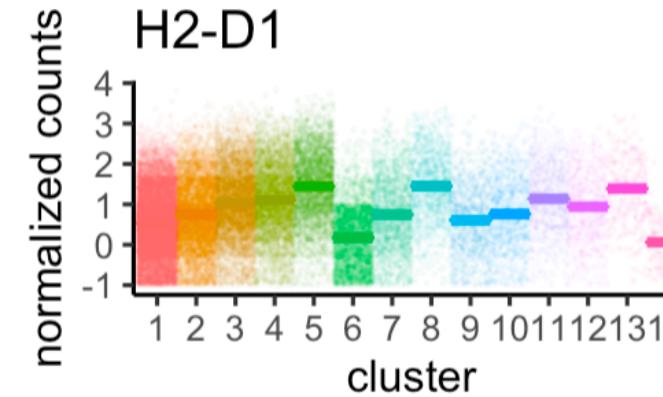
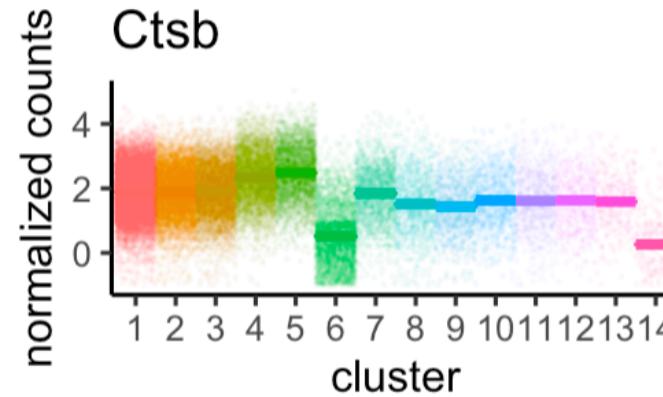
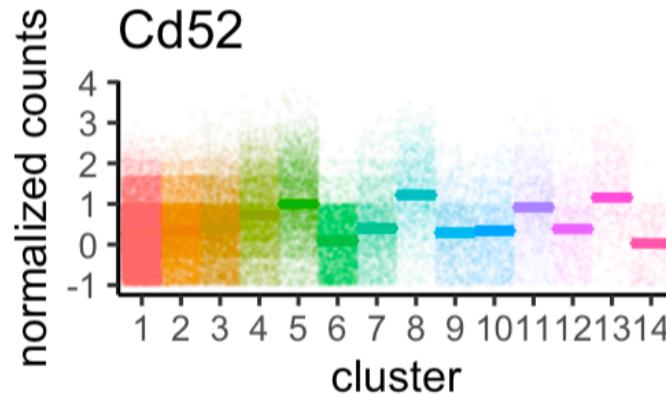
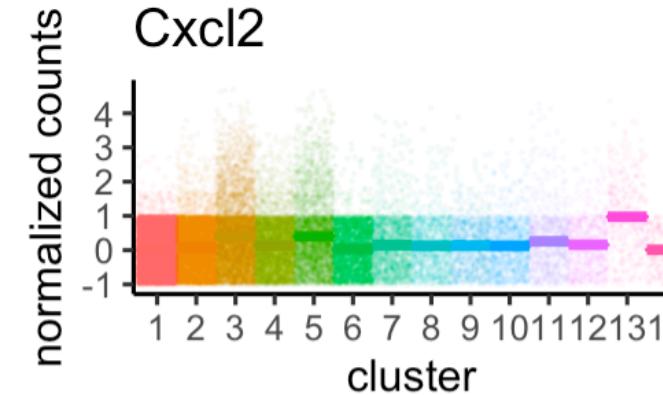
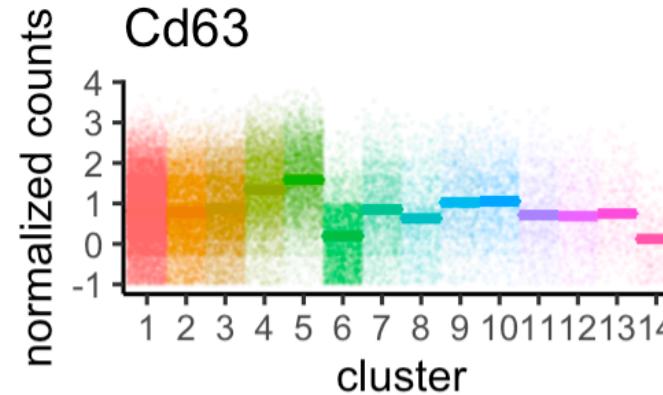
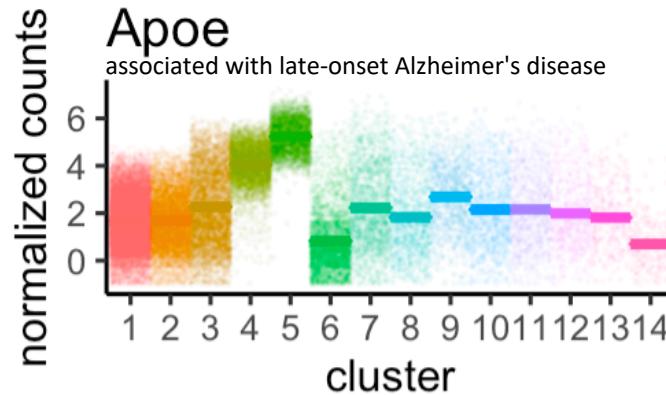
Induction of a common microglia gene expression signature by aging and neurodegenerative conditions: a co-expression meta-analysis

Inge R Holtman, Divya D Raj, Jeremy A Miller, Wandert Schaafsma, Zhuoran Yin, Nieske Brouwer, Paul D Wes, Thomas Möller, Marie Orre, Willem Kamphuis, Elly M Hol, Erik W G M Boddeke & Bart J L Eggen



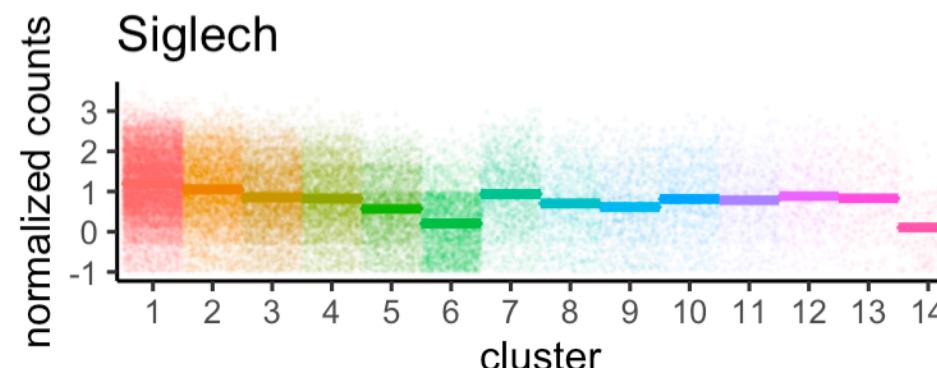
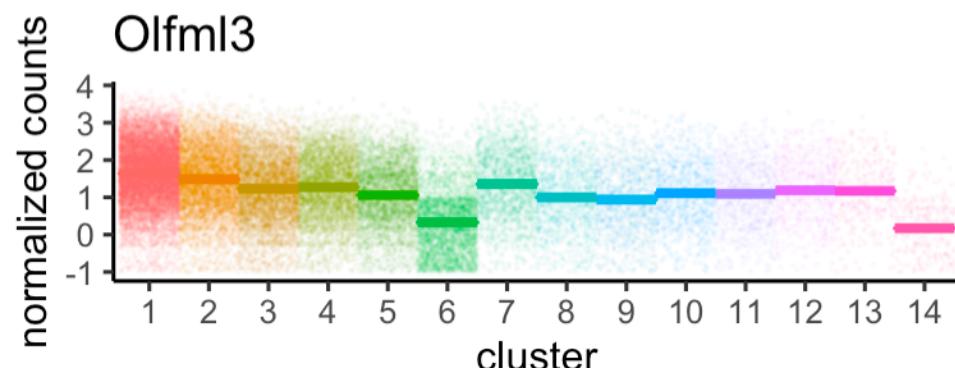
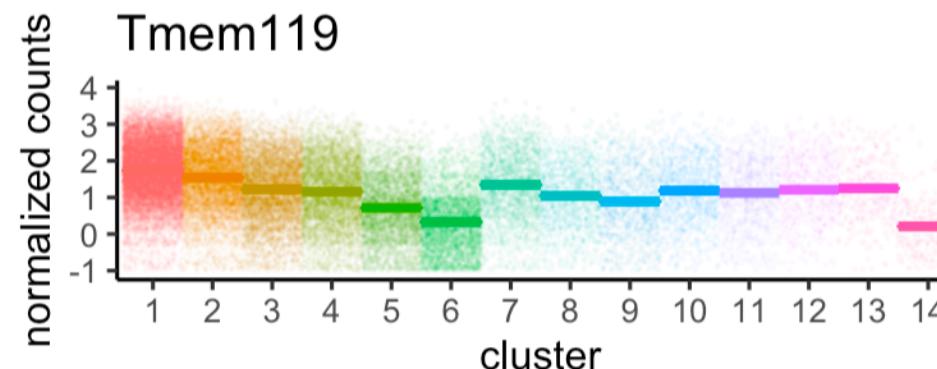
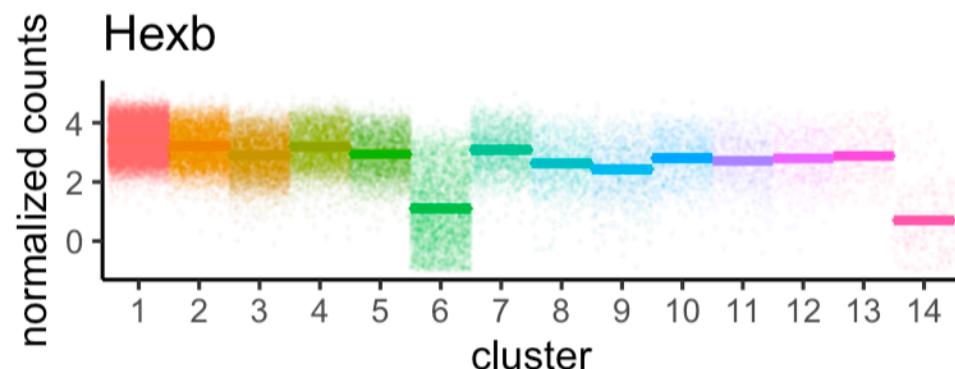
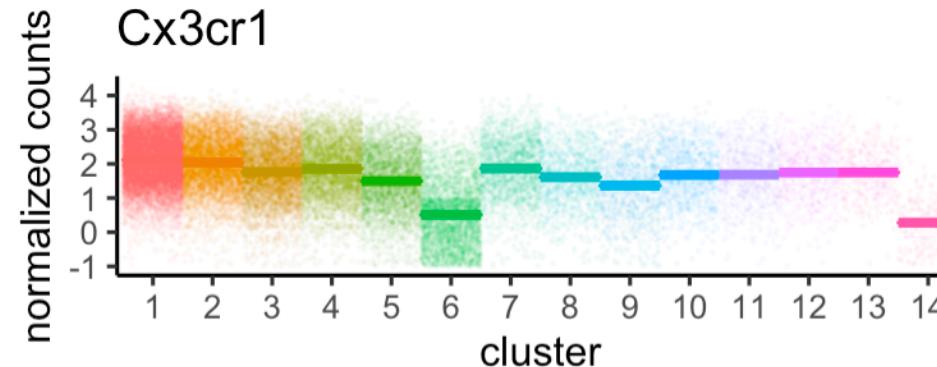
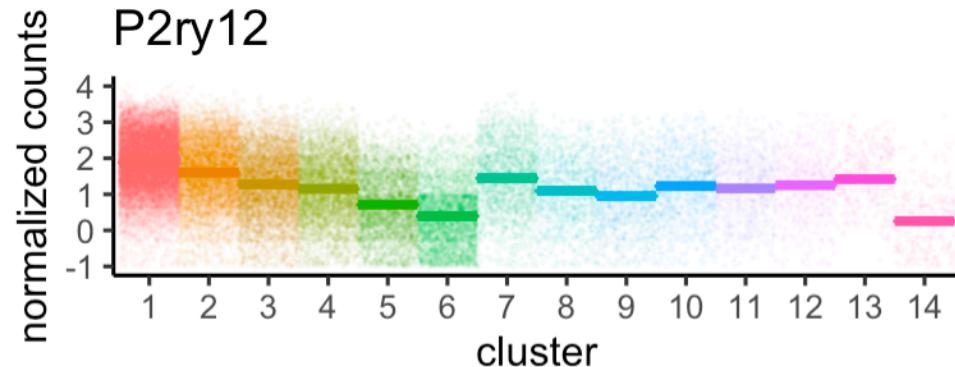
Top markers in both Cluster 5 and Arf1 include some of the well-known neurodegenerative disease risk genes



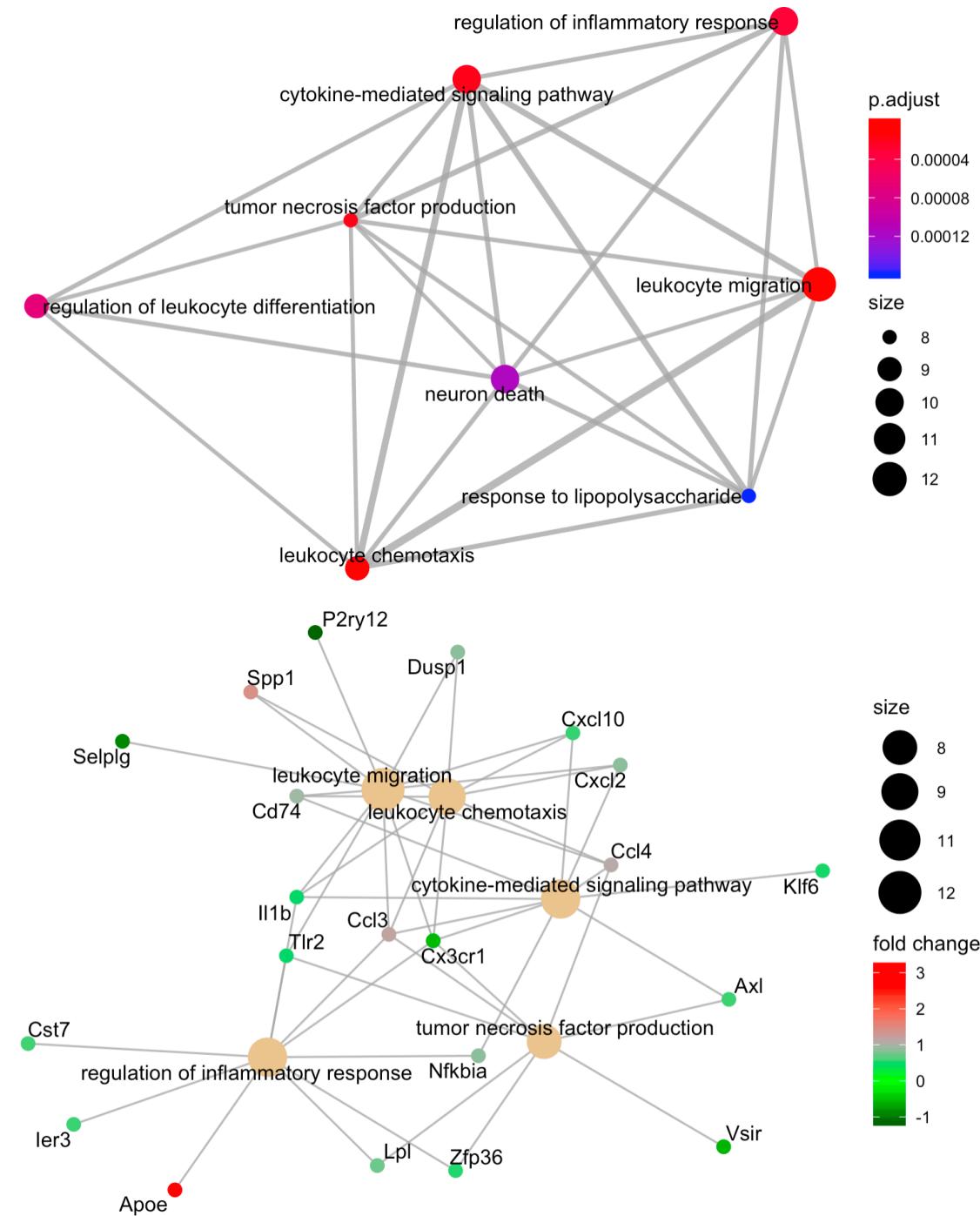
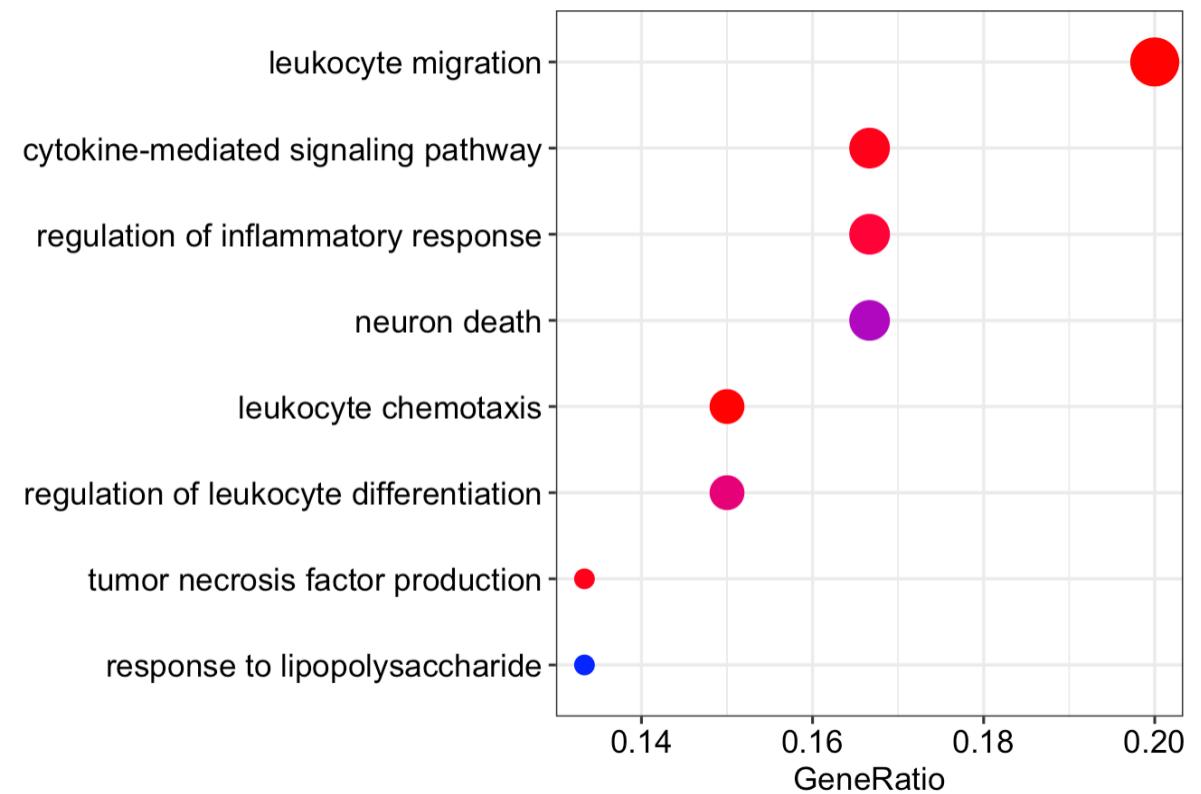


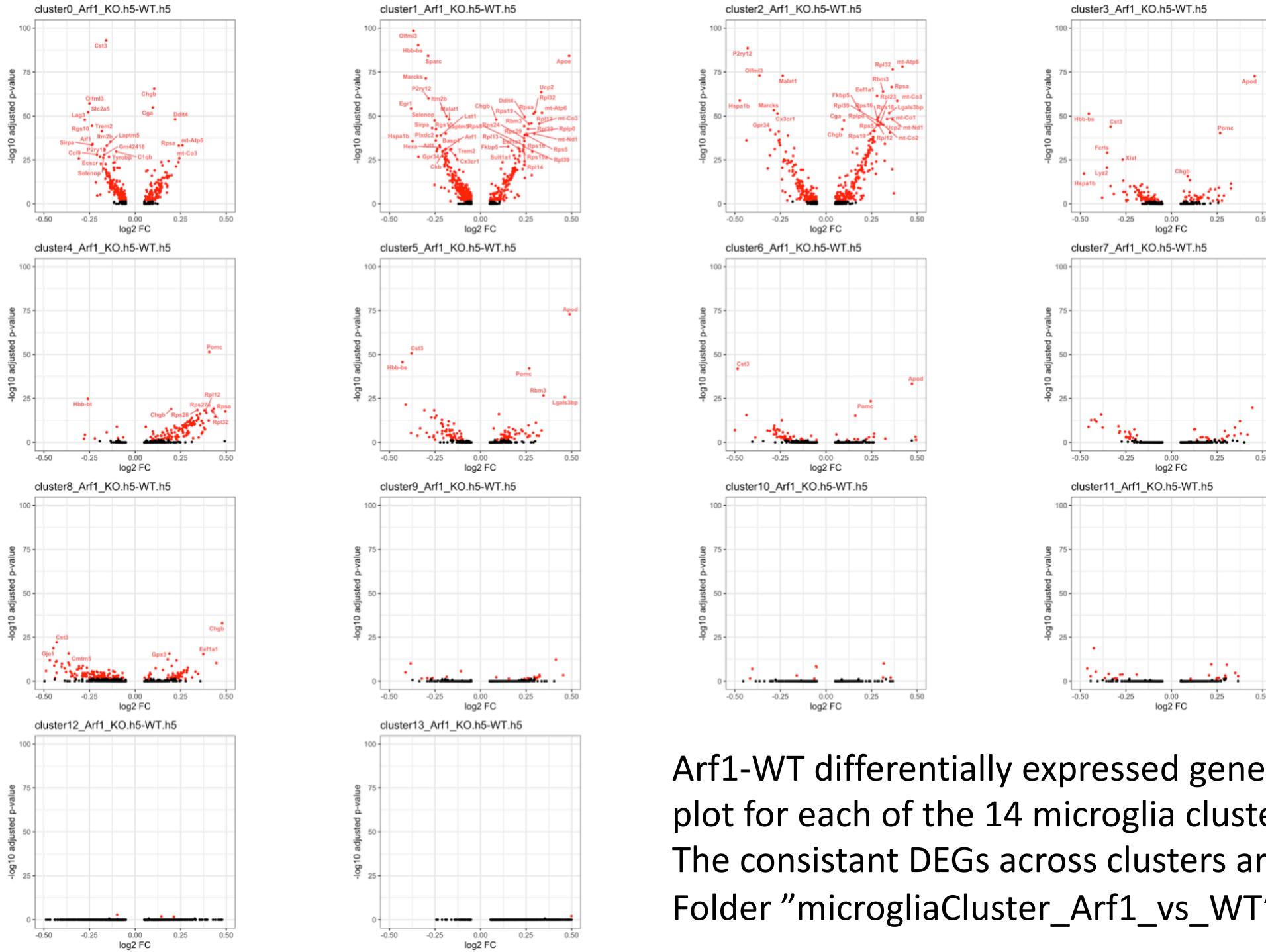
Top markers of Cluster 5 include many known disease risk factors

## Top marker of Cluster 1 include mostly “microglial marker genes”



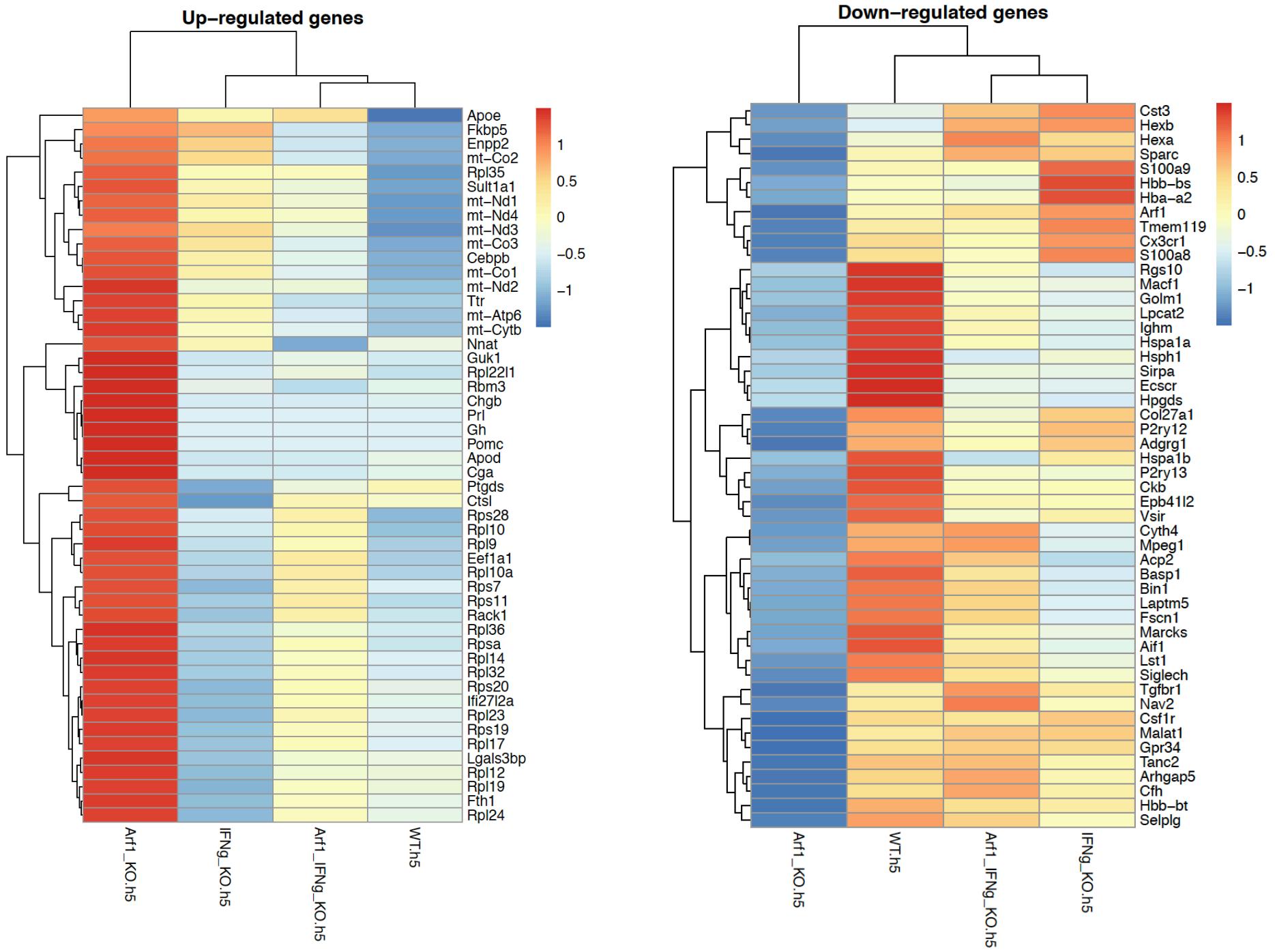
## Cluster5 - enriched pathways



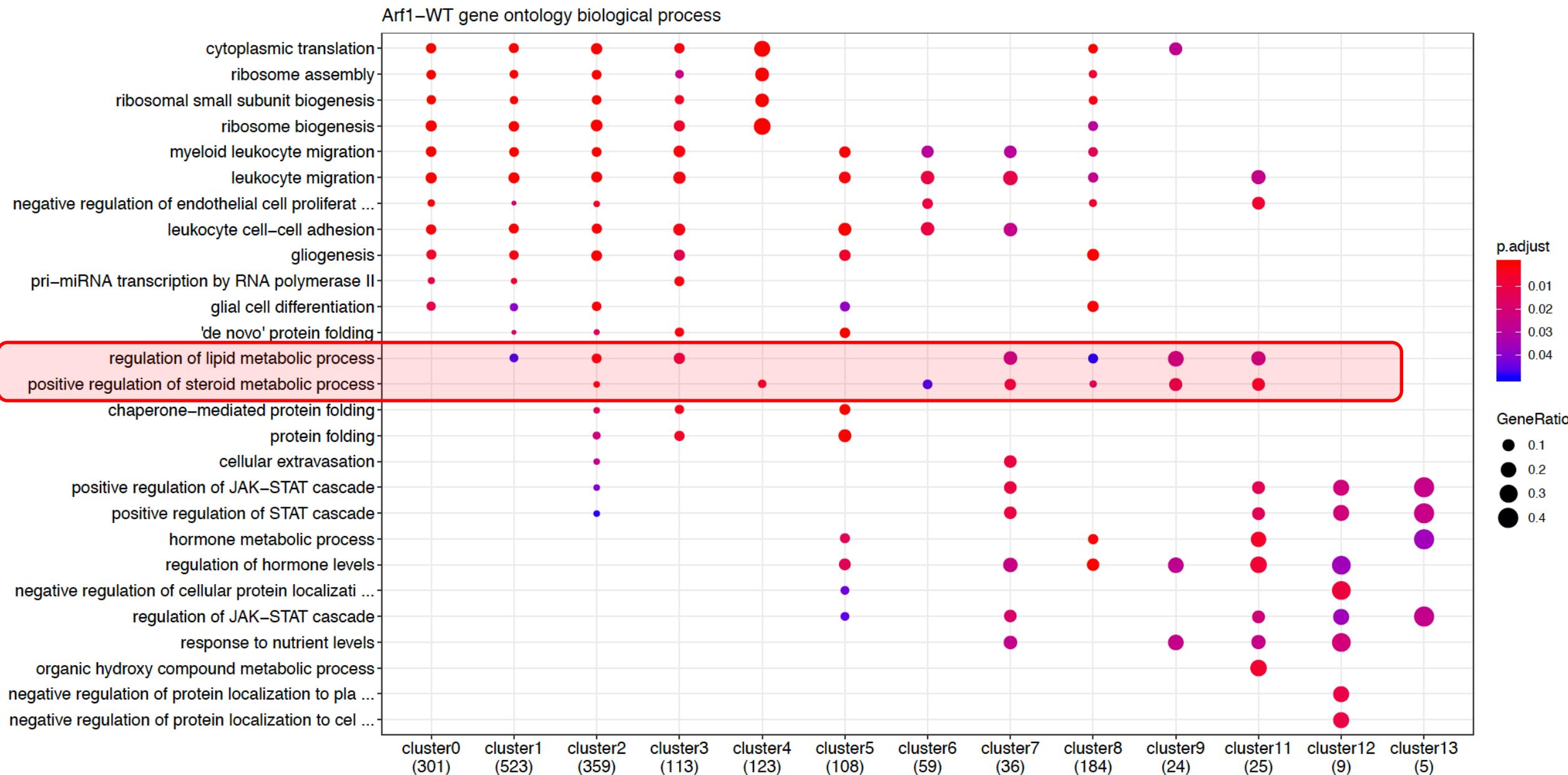


Arf1-WT differentially expressed genes volcano plot for each of the 14 microglia clusters  
The consistant DEGs across clusters are in  
Folder "microgliaCluster\_Arf1\_vs\_WT"

## Up and down-regulated genes in Arf1 vs. all



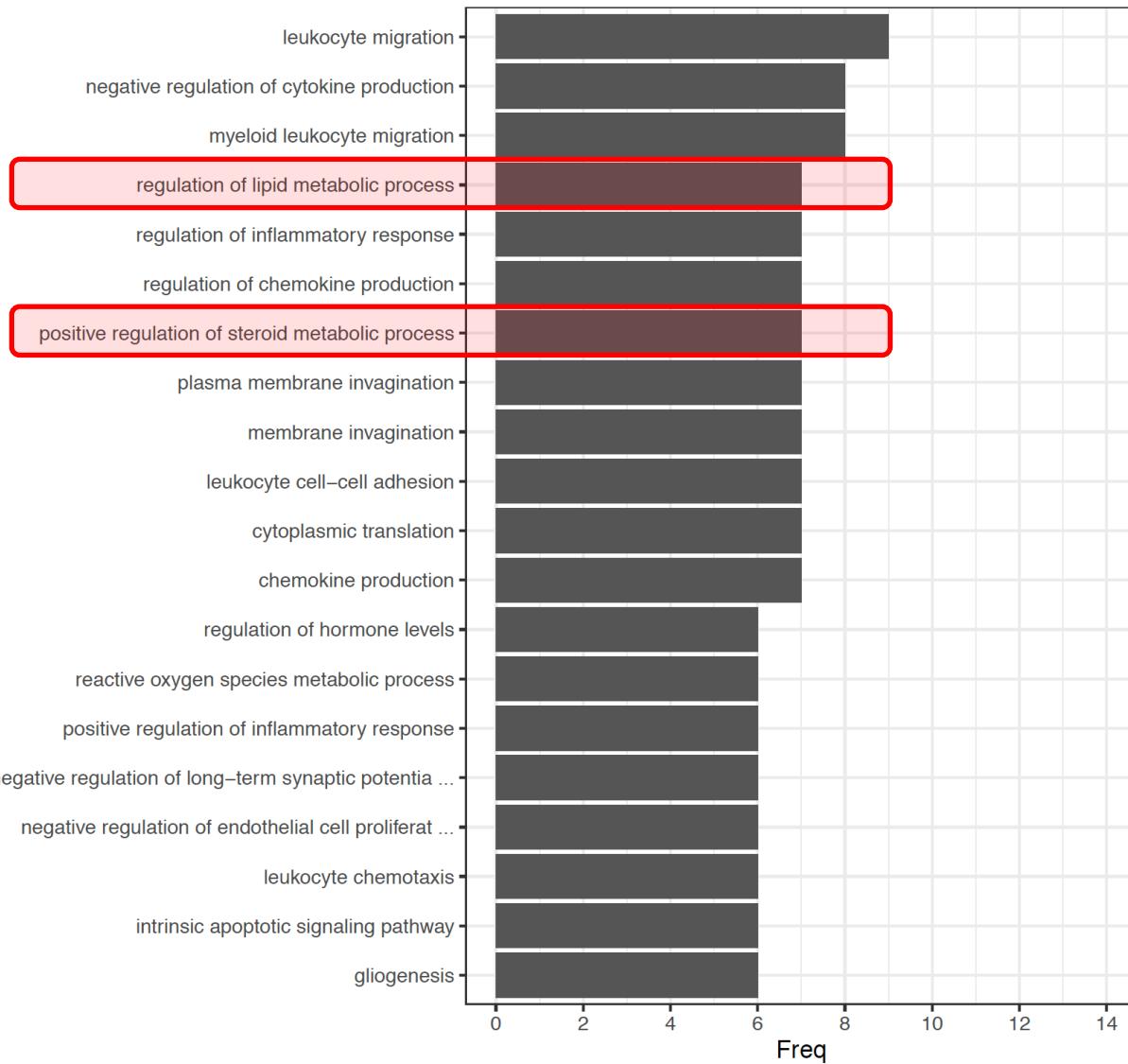
### Overlap of significantly over-represented processes in Arf1-WT for each microglial cluster



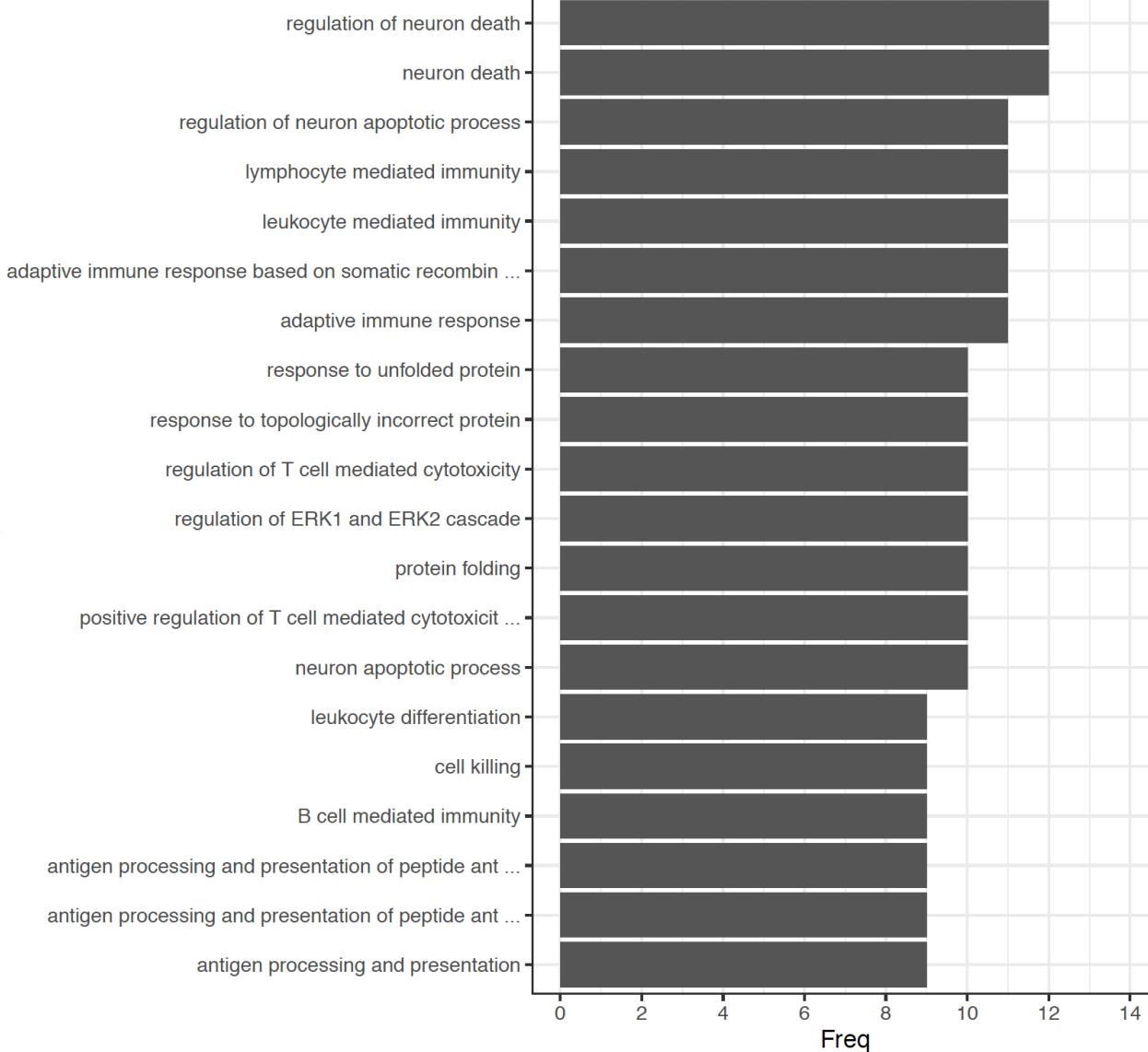
# Most frequent biological processes in different sample comparisons

biological process

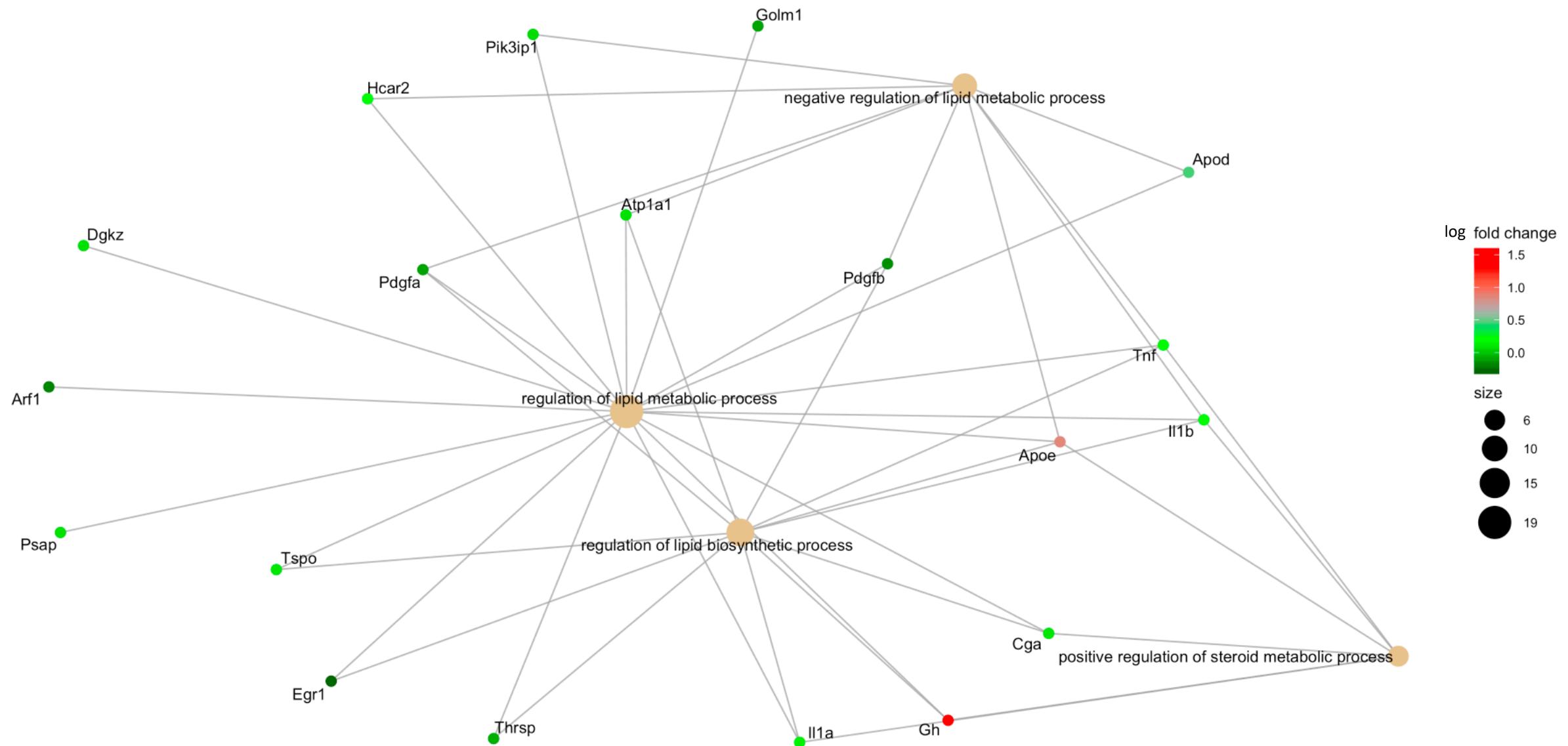
Arf1-WT : top enriched biological process



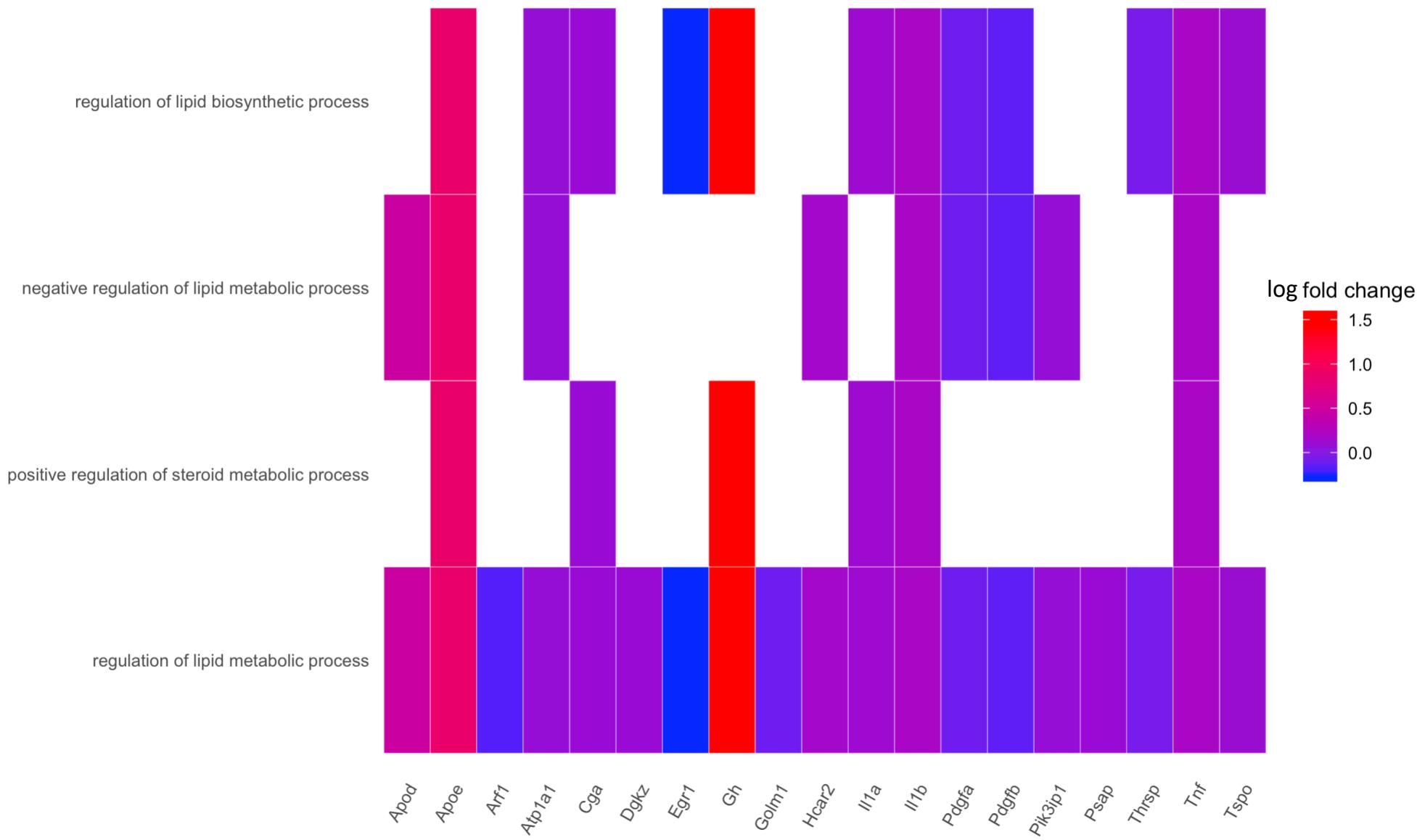
Arf1\_IFNg-WT : top enriched biological process



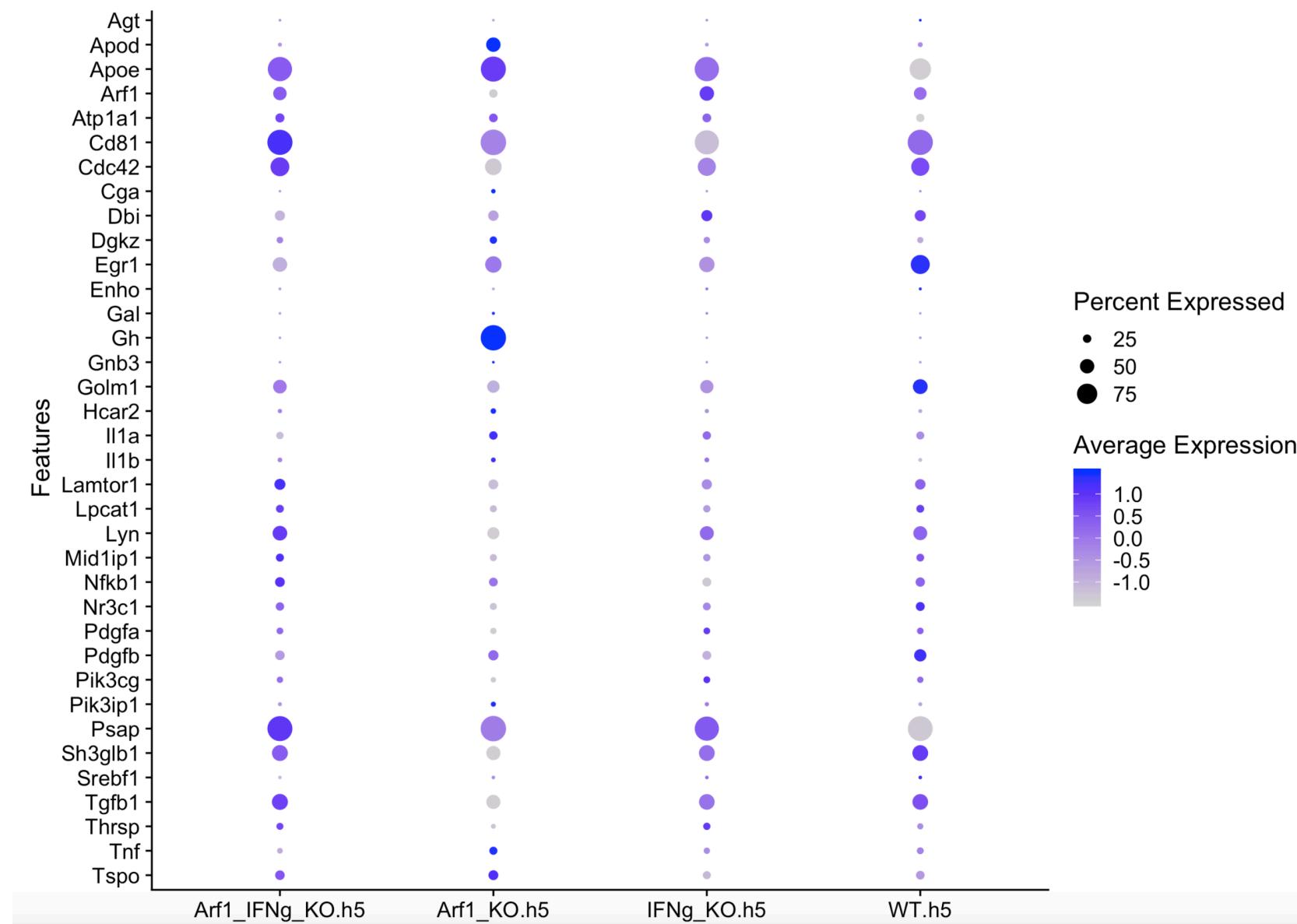
# lipid metabolic process network



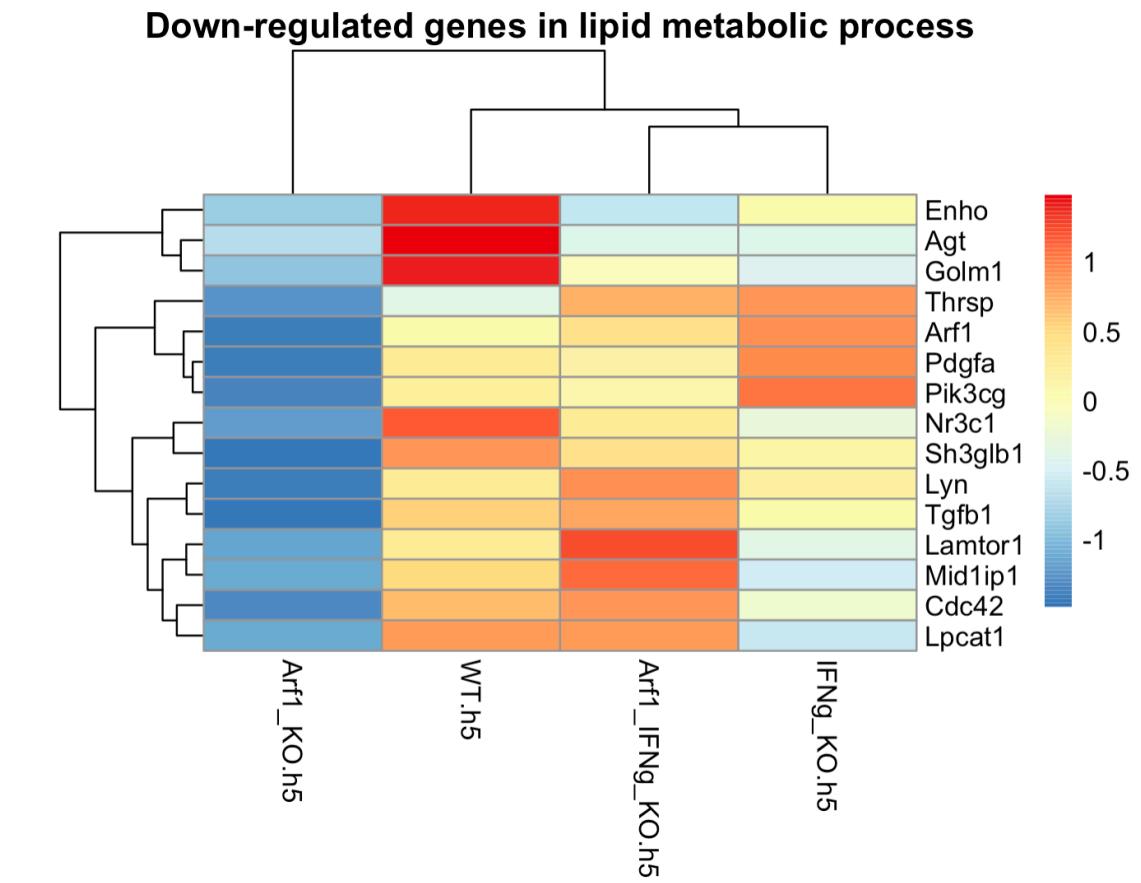
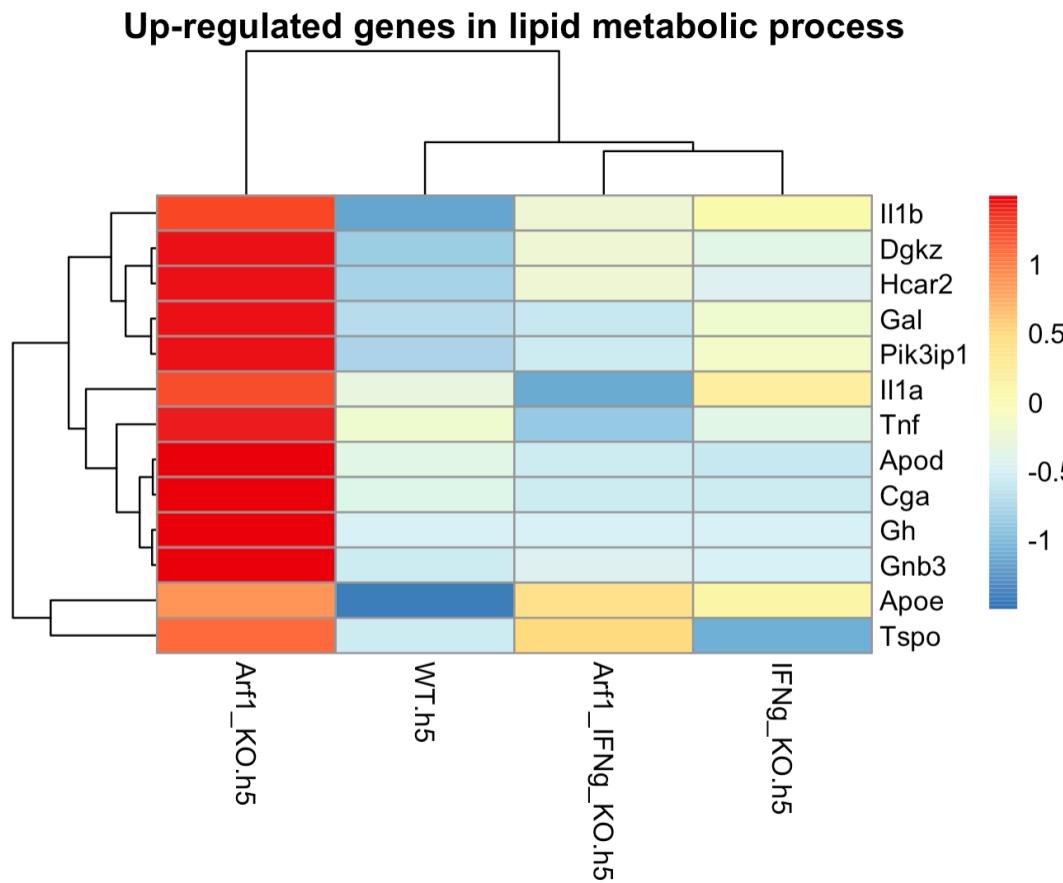
## Differentially expressed genes in the lipid metabolic process



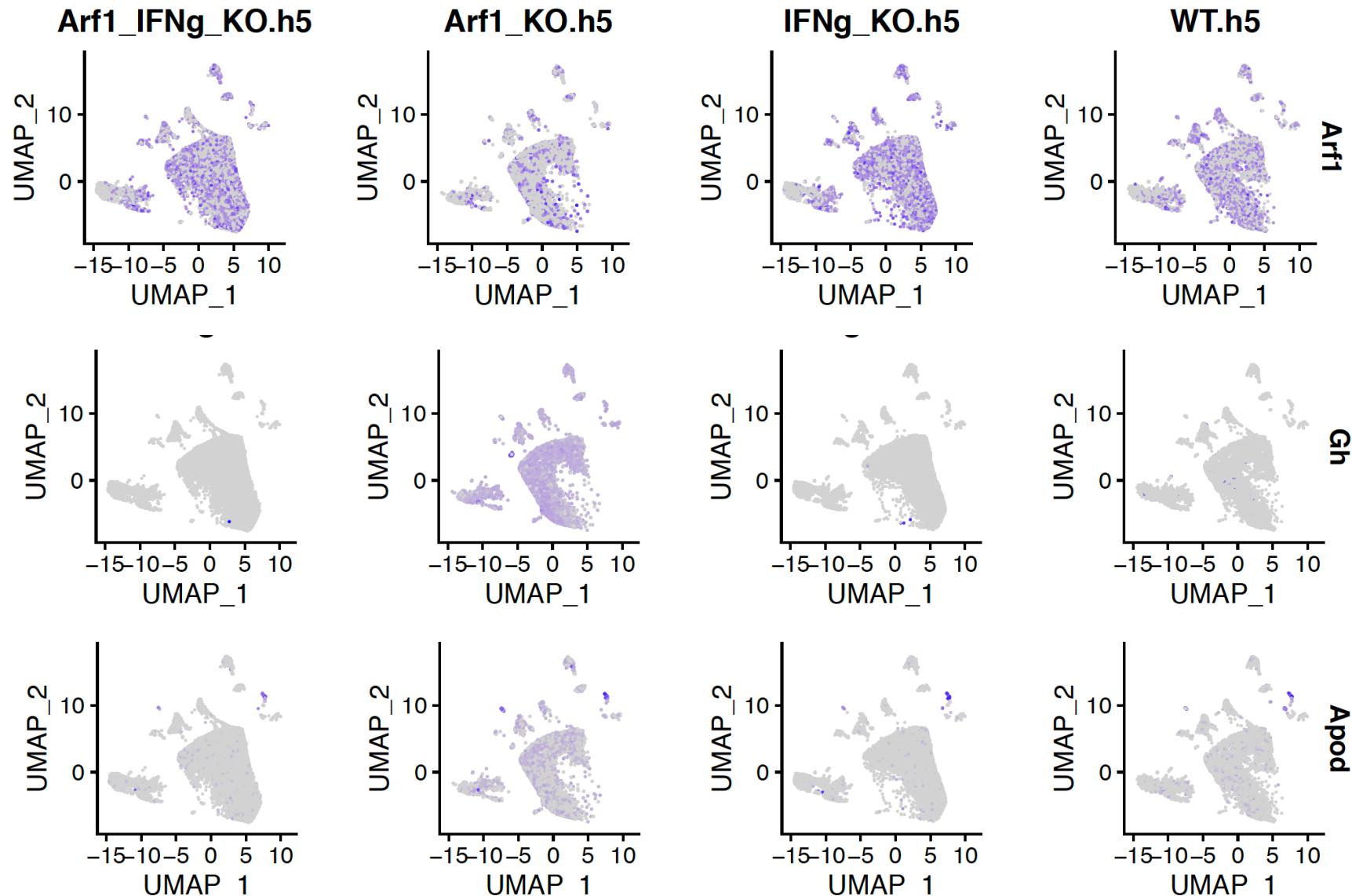
# Differentially expressed genes in the lipid metabolic process – across samples

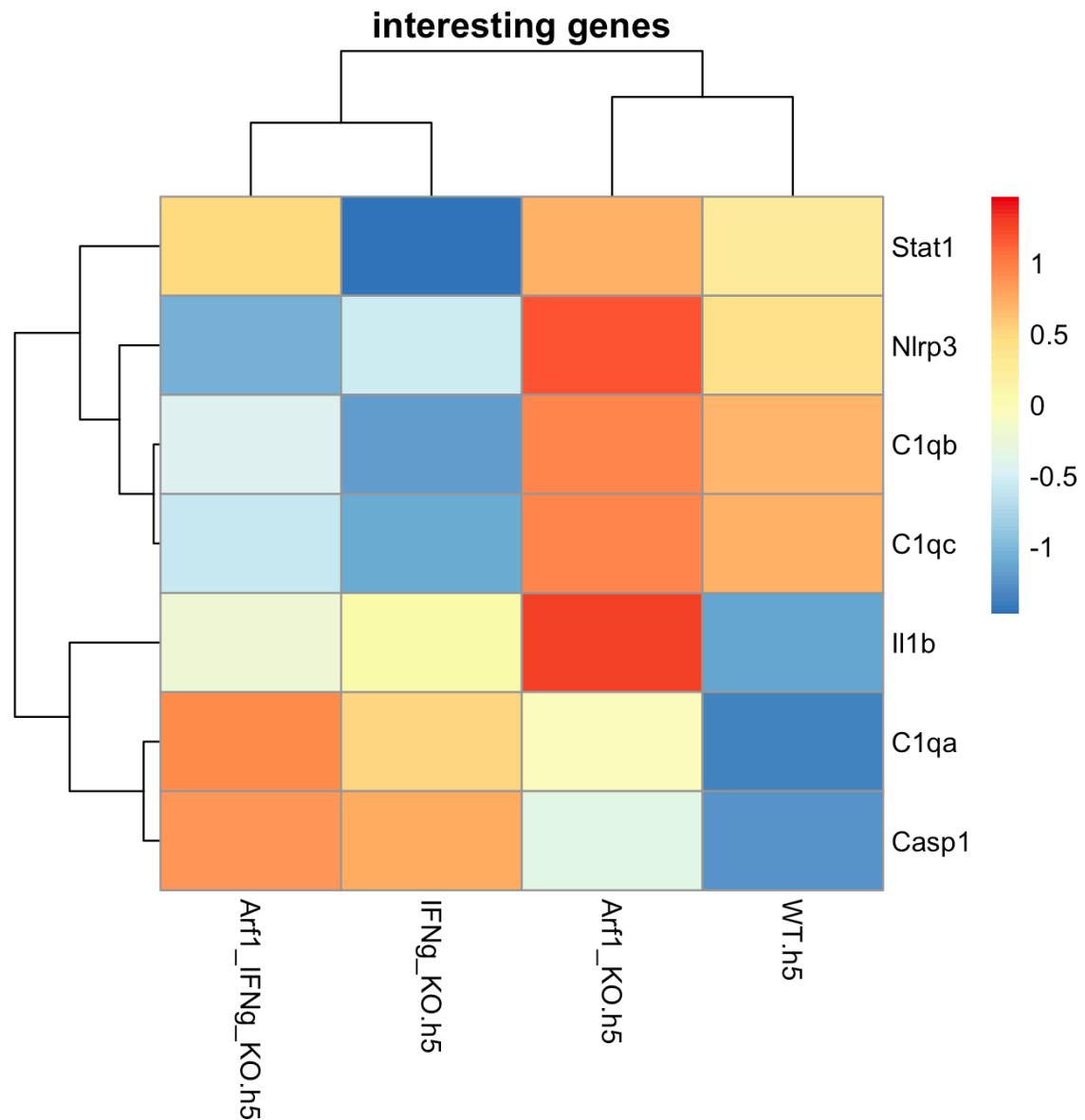


# Differentially expressed genes in the lipid metabolic process – across samples



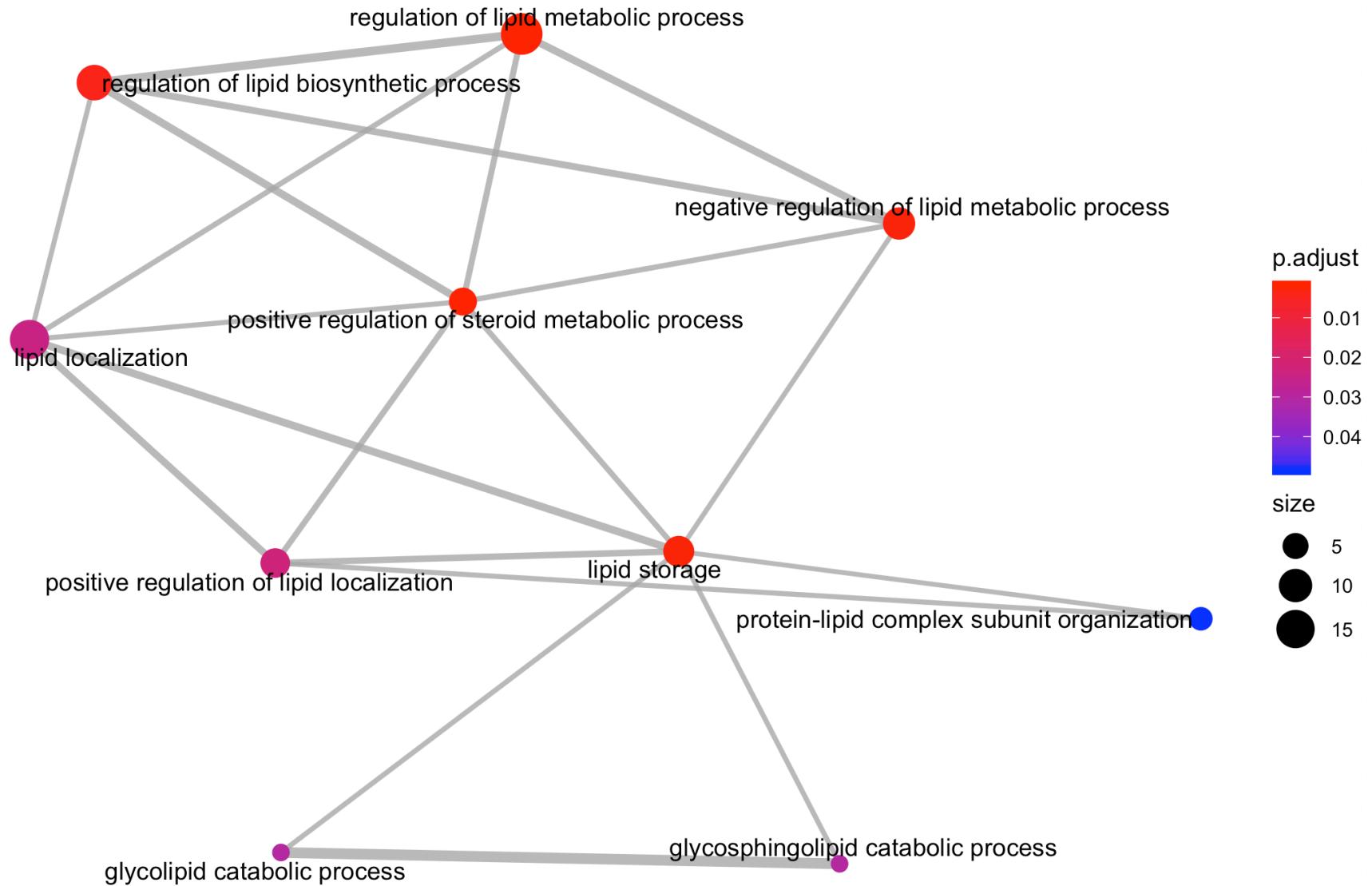
## Examples of down-regulated and up-regulated genes in the lipid metabolic process





Only "IL1b", "C1qa", "C1qb", "C1qc", "Stat1", "Nlrp3", "Casp1" are in microgliaReshape  
 "TNF" and "IL1a" are not in microgliaReshape

Enrichment map - enriched terms are organized into a network with edges connecting overlapping gene sets. In this way, mutually overlapping gene sets tend to cluster together, making it easy to identify functional module.



- A common feature of neurodegenerative diseases is the presence of activated microglia in areas of neuronal death.
- Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease characterized by progressive loss of motor neurons, leading to paralysis and death. Microglia activation occurs robustly in ALS patient tissue and in spinal cords of mutant SOD1 transgenic mice.

For comparisons of transcriptome changes in LDAM with published datasets (Fig. 2g; Extended Data Fig. 3), we selected the following published RNA-seq datasets of microglia in aging and neurodegeneration (modified after Bohlen et al., 2017)<sup>56</sup>: 24-month versus 4-month-old wild-type mice for aging (Holtman et al., 2015)<sup>22</sup>; APP+ versus APP- for Alzheimer's disease (Wang et al., 2015)<sup>24</sup>; SODG93A endstage versus nontransgenic day 130 for ALS (Chiu et al., 2013)<sup>23</sup>; DAM versus homeostatic microglia (Keren-Shaul et al. 2017)<sup>3</sup>; MGnD versus homeostatic microglia (Kraseman et al., 2017)<sup>4</sup>; and microglia clusters published by Li et al., (2019)<sup>5</sup> and Hammond and colleagues (2019)<sup>25</sup>. For gene set comparisons, we generated lists for the published datasets by using a fold change cutoff (modified after Bohlen et al., 2017)<sup>56</sup> and compared these lists with the top 100 upregulated and downregulated genes in LDAM.

## Arf1-WT

